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:

Ridge Regression:-

When we plot the curve between negative mean absolute error and alpha we see that as the value of alpha increase from 0 the error term decrease and the train error is showing increasing trend when value of alpha increases. when the value of alpha is 30 the test error is minimum so we decided to go with value of alpha equal to 20 for our ridge regression.

For lasso regression I have decided to keep very small value that is 50, when we increase the value of alpha the model try to penalize more and try to make most of the coefficient value zero. Initially it came as 0.4 in negative mean absolute error and alpha.

The most important variable after the changes has been implemented for ridge regression are as follows:-

- MSZoning
- Neighborhood
- SaleCondition
- GrLivArea
- Exterior1st

The most important variable after the changes has been implemented for lasso regression are as follows:-

1. GrLivArea

2. OverallQual

3. OverallCond

4. TotalBsmtSF 5. BsmtFinSF1 6. GarageArea 7. Fireplaces 8. LotArea 9. LotArea 10. LotFrontage

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 :

Both Ridge and Lasso can see that the r2_scores are almost same for both of them but as lasso will penalise more on the dataset and can also help in feature elimination. So we can consider lasso 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?

Five values that best describe the final model are as follows:

MiscVal : Value of miscellaneous feature

BsmtHalfBath : Basement half bathrooms

LowQualFinSF : Low quality finished square feet (all floors)

BsmtFullBath : Basement full bathrooms

HalfBath : Half baths above grade

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: Making the model more generalisable may take a toll on accuracy up to some extent but we can also have a look at the precision and recall of the model because sensitivity and specificity also play an important role in the model evaluation criteria. Together if all three are above average we may accept the model. A very accurate model may have a chance of getting overfitted.