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 for ridge regression = 100

The optimal value for lasso regression = 0.001

When I double the value of alpha in ridge regression , which is 200, then my R square on train and test data reduces. Also mean square error on train and test data increases. Most important predictor variable is “GrLivArea” with coefficient value of 0.062

When I double the value of alpha in ridge regression , which is 200, then my R square on train and test data reduces. Also mean square error on train and test data increases. Most important predictor variable is “GrLivArea” with coefficient value of 0.130

Same is observed when I double the alpha value of lasso regression to 0.002.

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:

I will use Lasso. Reason below

The primary difference between Lasso and Ridge regression is their penalty term. The penalty term here is the sum of the absolute values of all the coefficients present in the model. With Lasso, the penalty pushes some of the coefficient estimates to be exactly 0, provided the tuning parameter, λ, is large enough. Lasso performs feature selection

This variable selection results in models that are easier to interpret.

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:

Those 5 most important predictor variables that will be excluded are :-

1. GrLivArea

2. OverallQual

3. OverallCond

4. TotalBsmtSF

5. SaleType_New

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 robust and generalisable the model should be simple. Complexity should not be high. We can use the bias and variance scale to get a robust model

What we need is lowest total error, i.e., low bias and low variance, such that the model identifies all the patterns that it should and is also able to perform well with unseen data.

For this, we need to manage model complexity: It should neither be too high, which would lead to overfitting, nor too low, which would lead to a model with high bias (a biased model) that does not even identify necessary patterns in the data.

Accuracy should not vary if we have the model with low bias and low variance.