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 0.0001 and for Lasso it is 2.

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around the same but there is a small change in the coefficient values.

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 – 0.0001
- Lasso – 2

The Mean Squared Error of both the models are almost the same.

As Lasso contributes to feature reduction by setting the coefficient values of certain features to zero, it holds a distinct advantage over Ridge. Therefore, Lasso is recommended as the preferred choice for the final 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?

Answer:

The five most important predictor variables in the model are:-

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

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:

According to Occam's Razor, when faced with two models exhibiting similar performance in finite training or test data, the preference should be for the one that makes fewer assumptions on the test data. This is attributed to several reasons:

- Simpler models are generally more versatile and applicable across a broader range of scenarios.
- They demand fewer training samples for effective training compared to more complex models, making them easier to train.
- Simpler models tend to be more robust. Complex models can exhibit erratic changes with variations in the training dataset.

- Simple models, characterised by low variance and high bias, often outperform complex models, which tend to have low bias and high variance. However, simpler models may make more errors in the training set.

Complex models, on the other hand, may lead to overfitting — performing exceptionally well on training samples but failing when applied to other test samples. To enhance model robustness and generalizability, the recommendation is to keep the model simple without making it overly simplistic, as excessive simplicity may render it ineffective.

Regularisation serves as a tool to achieve this delicate balance, helping to make the model simpler without compromising its utility. In regression, regularisation entails introducing a regularisation term to the cost function, incorporating either the absolute values or the squares of the model parameters.