Advanced Regression Assignment

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer: Optimal Value of alpha for ridge and lasso regression are:

• Optimal Value of lambda for Ridge: 10
• Optimal Value of lambda for Lasso: 0.001

If we choose to double the value of alpha for both ridge and lasso:
In case of ridge that will lower the coefficients and in case of Lasso there would be more less important features coefficients turning 0.
The most important predictor variable after the change is implemented are those which are significant.

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:
Optimal Value of alpha for ridge and lasso regression are:
- Optimal Value of lambda for ridge: 10
- Optimal Value of lambda for Lasso: 0.001

As we Got good score for both the models so we can go with Lasso regression as it results in model parameters such that lesser important features coefficients become zero.

Ridge: Train :90.9 Test :87.4 and Lasso : Train :89.8 Test :86.4

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: On running the same notebook and removing the top 5 significant variables :
We found below variables as next 5 significant.

-Lasso

```
('GarageType_BuiltIn', 0.089),
('GarageType_Detchd', 0.094'),
('GarageType_No Garage', 0.101),
('GarageType_Others', 0.12),
('GarageFinish_No Garage', 0.195)]
```

-Ridge

```
('GarageType_builtIn', 0.089),
('GarageType_Detchd', 0.093'),
('GarageType_No Garage', 0.096),
('GarageType_Others', 0.103),
('GarageFinish_No Garage', 0.14)]
```

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:

- A model is **robust** when any variation in the data does not affect its performance much.
- A **generalizable** model can adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- To make sure a model is robust and generalizable, we must **take care it doesn't overfit**. This is because an overfitting model has very high variance and the smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data but fail to pick up the patterns in unseen test data.
- In other words, the model should not be too complex in order to be robust and generalizable.

- If we look at it from the perspective of **Accuracy**, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.
- In general, we must find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.