Advanced Regression Assignment
(Subjective questions –Part II)

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: The ideal hyperparameter values for Ridge and Lasso regression are as follows:

Optimal λ for Ridge: 10

Optimal λ for Lasso: 0.0006

If we decide to double the value of alpha for both Ridge and Lasso:

For Ridge regression, this will result in a reduction of coefficients, emphasizing a stronger regularization effect.

In the case of Lasso regression, doubling the value of lambda will lead to more coefficients associated with less important features being set to zero. This intensification of regularization enhances the sparsity of the model.

Following this adjustment, the predictor variables that remain significant and influential will be considered the most important.

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 best hyperparameter values for Ridge and Lasso regression are as follows:

Optimal λ for Ridge: 10

Optimal λ for Lasso: 0.001

Since both models achieved satisfactory scores, opting for Lasso regression is advisable. Lasso's advantage lies in its ability to set coefficients associated with less important features to zero, resulting in a more parsimonious model.

Model performance scores:

Ridge: Training score: 90.9, Test score: 87.4

Lasso: Training score: 89.8, Test score: 86.4

Question 3: *After building the model, you realized 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', 0.053),
 ('GarageType_Detchd', 0.057),
 ('GarageType_No Garage', 0.06),
 ('GarageType_Others', 0.061),
 ('GarageFinish_No Garage', 0.079),
```

-Ridge

```
('GarageType_BuiltIn', 0.053),
 ('GarageType_Detchd', 0.056),
 ('GarageType_No Garage', 0.058),
 ('GarageType_Others', 0.059),
 ('GarageFinish_No Garage', 0.072),
```

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 : Consider implementing the following modifications to enhance your model:

1. Model-related changes:
    - Utilize a model resilient to outliers. Tree-based models are generally less affected by outliers compared to regression-based models. If your analysis involves a statistical test, consider opting for a non-parametric test rather than a parametric one.
    - Adopt a more robust error metric. Transitioning from mean squared error to mean absolute difference (or an alternative like Huber Loss) can mitigate the impact of outliers. Further insights into this choice can be found in the explanation provided at "Why is the median a measure of central tendency? It doesn't have anything to do with any other values of the data set, so how does it 'describe' the data set?"
2. Data-related changes:
    - Transform your data. If your data exhibits a distinct right tail, consider applying a log transformation to alleviate skewness.
    - Address outliers. Removing outliers can be effective if they are few in number and can be confidently identified as anomalies with minimal predictive value.
    - Implementing these adjustments can contribute to a more robust and accurate modeling process.