

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 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 we will see the following changes.

- Ridge: It will lower the coefficients
- 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: The 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

In this case we got a good score for both the models, we can opt to go with Lasso regression as it results in model parameters such that lesser important features coefficients become zero.

Ridge: Training: 90.9, Test: 87.4

Lasso: Training: 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: After running the same notebook and removing the top 5 significant variables we found the following 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 generalisable? What are the implications of the same for the accuracy of the model and why?

Answer: Some changes that you can make to your model is as follows:

1. Use a model that is resistant to outliers- Tree-based models are generally not affected by outliers as compared to regression based models. While performing statistical test, try non-parametric test.
2. Use a more robust error metric- Switch from mean squared error to mean absolute difference (or Huber Loss) to reduce the influence of outliers.

Some changes you can make to your data:

- **Winsorize your data**

- **Transform your data**
- **Remove the outliers**