

## 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 1:

Optimal Value of Alpha:

- The optimal value of alpha in case of Lasso Regression (Primary Model): 0.001
- The optimal value of alpha in case of Ridge Regression (Primary Model): 8

Ridge Regression with doubled alpha (16):

- Ridge regression model with doubled alpha (16) lacks behind in terms of test accuracy when compared to the primary model (8).
- Upon comparing the mean squared error (MSE) test scores of the primary model (alpha = 8) and doubled alpha model (alpha =16), it was observed that the primary model has a lower MSE than the doubled alpha model.
- The performance of the primary model is better as compared to the doubled alpha model on the test and train set.

Lasso Regression with doubled alpha (0.002):

- The test accuracy of the lasso regression model with an alpha value of 0.001 is slightly higher compared to the doubled alpha model with an alpha value of 0.002.
- The MSE test scores of similar data from the original dataset and the doubled alpha model show that the single alpha model has slightly smaller scores than the doubled alpha model.
- In terms of train and test data, the single alpha Lasso Regression model performs better than the doubled alpha model.

## 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 2:

Optimal Values of Alpha:

- Ridge Regression- 8.0
- Lasso Regression-0.001

The Lasso Regression Model has a better R2 test score, lower MSE, better performance on the test set.

Feature selection in Lasso regression is also easier as the coefficient values of some insignificant predictor variables become 0.

### **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 3:**

Top 5 features after dropping 5 main predictors–

- 1stFlrSF
- MSSubClass\_90
- MSSubClass\_120
- TotalBsmtSF
- HouseStyle\_1Story

### **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 model robust and generalisable 3 features are required:

- Model accuracy should be > 70-75%: In our case its coming 80%(Train) and 81%(Test) which is correct.
- P-value of all the features is < 0.05
- VIF of all the features are < 5

Thus, we are sure that model is robust and generalisable.