

## Problem Statement - 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 when the change is implemented?

Ans:

The optimal value of alpha for Ridge is found to be 0.005, and for Lasso, it is 19.

In Ridge regression, as alpha increases, the coefficient values increase. The R<sup>2</sup> score of the training data drops from 0.807 to 0.45. For Lasso, an increase in alpha leads to the removal of more features from the model, resulting in a 1% drop in R<sup>2</sup> scores for both test and train data.

### 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?

Ans:

Lasso regression will be chosen due to its feature selection capability. It effectively removes unnecessary features from the model without compromising accuracy, making the model more generalized, simpler, and accurate.

### 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?

Ans:

### 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?

Ans:

To ensure model robustness and generalizability, three criteria must be met:

The model accuracy should be greater than 70-75%. In this case, the accuracy is 87% for training and 81% for testing, meeting the required threshold.

All feature p-values should be less than 0.05.

The Variance Inflation Factor (VIF) for all features should be less than 5.

Meeting these criteria confirms the model's robustness and generalizability, contributing to its reliability in predicting outcomes.

