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?

Answered In code

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?

Answered in code

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?

Answered in code

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 the model robust and generalisable: more training data should be used so that it takes into account the variance, model with its parameters should be simple and coefficients should not be large. To make the model simpler, regularization techniques like Lasso and Ridge are applied. They penalize the coefficients of the model and do not let them scale up. We know that the model is generalized and robust when we get similar accuracy for test and train data. This is more desirable than a model which fits maximum variance of the training data (overfitting) and gives a high accuracy on just test data. The model should be generalizable so that it is not much affected by the outliers. This implies that instead of learning the data points(along with noise), the model has learned the essence of the trend in data. And hence is not much affected by the variance in training data.