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:

Optimum alpha for ridge = 20

Optimum alpha for lasso = 100

Change in model with double the value of alpha in ridge:

      The value of AIC and BIC increased and adjusted R2 decreased slightly.

Change in model with double the value of alpha in lasso:

      The value of AIC and BIC increased and adjusted R2 decreased slightly.

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:

I will choose to apply lasso regression as it has eliminated the variables which are not significant. The value of AIC and BIC are lowest and adjusted R2 is highest among the model for lasso.

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?

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?

Answer: By using regularization we can make model robust and generalizable. Regularization remove multicollinearity and keep check on overfitting or underfitting thus the model works well with the unforeseen data as well. This is how it make the model more generic.