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

The optimal value of alpha for ridge and lasso regression can be determined from the code provided in the previous answer. The optimal alpha for Ridge regression is 100.0, and the optimal alpha for Lasso regression is 0.1.

If we double the value of alpha for both Ridge and Lasso regression, the model coefficients will be penalized more heavily, leading to a more constrained model. This means that the model will become simpler and may have fewer predictors. The most important predictor variables may change, and the ones with lower coefficients may become more significant.

For our case we have the optimal value of 0.0811130

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

The choice between Ridge and Lasso regression depends on the problem requirements. Ridge regression is better suited for problems where we have a large number of predictors, and all of them are expected to be relevant. Lasso regression is useful when we want a sparse model with only a few predictors.

In this case, we have not specified any specific requirements, so it's hard to say which method is better. However, since Lasso regression yields a sparse model with only a few predictors, it can be easier to interpret and more computationally efficient.

We have got 0.0811130 as the best case scenario

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

To create another model excluding the five most important predictor variables, we need to first identify those variables. We can do this by looking at the coefficients of the Lasso model and selecting the five variables with the lowest coefficients. The five most important predictor variables after the change is implemented will be the five variables with the highest coefficients in the new Lasso model.

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 4

To ensure that a model is robust and generalizable, we can follow these best practices:

1. Use a large, representative data set that covers a wide range of scenarios.
2. Split the data set into training and testing sets and use cross-validation to evaluate the model performance.
3. Use regularization to prevent overfitting and improve model generalization.
4. Choose a simple model that is easy to interpret and explain.

5. Perform sensitivity analysis to assess how the model performance changes with different parameter values.

The implications of making a model robust and generalizable are that the model may sacrifice some accuracy on the training set to achieve better performance on the testing set. This is because the model is designed to generalize well to new data and not just memorize the training set. However, a more robust and generalizable model will be more reliable and trustworthy in making predictions on new data.