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:

- The optimal value of the regularization parameter (alpha) in ridge and lasso regression is typically not a fixed value and depends on the specific dataset and the trade-off between fitting the model to the training data and preventing overfitting.

  so., we have three cases

  - $\alpha=0$ - Get the same coefficients for linear regression as before
  - $0<\alpha<\infty$ - The coefficients are between 00 and the normal ones for linear regression
  - $\alpha=\infty$ - All coefficients are 0

  The optimal value of alpha can be found through various techniques such as cross-validation, grid search, or AIC/BIC criteria. There is no generic "optimal" value of alpha, and the appropriate value must be determined through empirical testing.

  The choice of alpha should be informed by the trade-off between the bias-variance trade-off in the model and the magnitude of the regularization penalty. However, in general, increasing alpha tends to lead to a simpler model with fewer features having a significant impact.

- The changes in the model if you choose double the value of alpha for both ridge and lasso the regularization effect would be stronger. The model would become more constrained, and the coefficients would be pushed further towards zero. For lasso, this could result in more coefficients being exactly zero.
  In Assignment,

- The important predictor variable is difficult to predict without knowing the details of the dataset and the initial values of alpha. In ridge regression, all variables generally contribute to some extent, but with lasso, some variables may be completely excluded (coefficients set to zero). The most important predictors would be those with non-zero coefficients after the change.
  In Assignment,

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:

The optimal value of lambda for ridge and lasso regression during the assignment is .

For which one will be chosen and apply, we might want to compare their performance based on metrics such as mean squared error (MSE) or cross-validated performance. Additionally, if interpretability and sparsity are important, lasso may be a preferred choice.

The choice between ridge and lasso can be subjective and may involve some trial and error. It's common to try different regularization techniques and parameters, evaluate their performance, and choose the one that best aligns with our modeling goals and the characteristics of our data.

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?

Answer:

The 5 most predictor values excluded are:

The lasso model has likely set the coefficients for these variables to zero, effectively excluding them from the model. The new model would be those with non-zero coefficients. The remaining variables with non-zero coefficients are the ones the model considers most important in predicting the target variable.

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?

To keep the model to be robust and generalizable, need to make sure the model is not overfit and keep it as simple as possible. The accuracy of the model will go up if we try to over fit the model but that no longer makes it generalizable.

When the model is generalized the accuracy should be pretty good on both the training and the testing dataset making the model robust. The focus on generalization helps to ensure that the model's accuracy extends beyond the specific data used for training, contributing to its overall effectiveness and reliability.