Q.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 :
Optimal value of alpha for ridge and lasso regression:

1. The optimal value of alpha depends on the specific dataset and the goal of the analysis. It is best to experiment with different alpha values and evaluate the model's performance using appropriate metrics (e.g., cross-validation, mean squared error, R-squared) to determine the optimal value.
2. Higher values of alpha result in stronger regularization, which can help reduce overfitting but may also lead to increased bias. Lower values of alpha reduce regularization, allowing the model to fit the training data more closely but increasing the risk of overfitting.
3. Higher regularization can lead to further reduction in the magnitude of coefficients, resulting in a simpler model with fewer significant predictors.

Q.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 : Ridge regression as it is suitable when there is a possibility of multicollinearity (high correlation) among the predictor variables. It can help reduce the impact of multicollinearity by shrinking the coefficients.

Q.3
Answer: Build a new Lasso regression model:

1. Use the same dataset and preprocessing steps as before, excluding the five predictor variables that are not available in the incoming data.
2. Set the regularization parameter (alpha) based on the optimal value.
3. Train the Lasso model
4. Retrieve the coefficients
5. Sort and select the top five predictor variables

Q.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 :

1. Train the model on a sufficiently large and representative dataset
2. Perform cross-validation, such as k-fold cross-validation, to evaluate the model's performance on multiple subsets of the data.
3. Apply regularization techniques like L1/L2 regularization (lasso/ridge) to mitigate overfitting and enhance generalizability.
4. Search over a range of hyperparameter values to find the combination
5. Assess the model's performance using appropriate evaluation metrics