Q1. 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?

Ans:- optimal value of ridge and lasso regression is typically determined through a process called hyperparameter tuning. The right alpha value balances the trade-off between the model's bias and variance, with a smaller alpha leading to a less regularized (closer to a standard linear regression) model and a larger alpha leading to a more regularized model.

Ridge regression (L2 Regularization) : alpha in Ridge regression are 0.1, 1, 10, or you can use a more granular search over a wider range.

The optimal alpha value minimized the MSE on validation data

Lasso regression (L1 Regularization) : Lasso can also perform feature selection as it can force some coefficients to be exactly zero. Therefore, it can eliminate irrelevant features.

increasing alpha is a form of feature selection and regularization, and the exact impact on predictor variables depends on the data and the original values of the coefficients. You would need to experiment with different alpha values and evaluate the model's performance to determine the effect on the most important predictors.

Q2: 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?

Ans:- Lasso (L1 Regularization): Use Lasso when you suspect that many of your predictor variables are irrelevant or redundant, and you want to perform feature selection by forcing some coefficients to be exactly zero.

If we have a high-dimensional dataset with many features, Lasso can be a powerful tool for feature selection and dimensionality reduction.

Ridge (L2 Regularization): the decision should be based on the nature of your data and the goals of your analysis. If you've already determined the optimal value of lambda for both Ridge and Lasso through cross-validation, consider the following:

optimal lambda for Lasso is very close to zero, it suggests that Lasso is essentially performing like standard linear regression and not forcing many coefficients to be exactly zero. In this case, Ridge might be a better choice to control multicollinearity and improve generalization.

Ridge and Lasso should depend on your specific needs regarding feature selection, multicollinearity, and the overall performance of your regression model.

Q3 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?

Ans:- five most important predictor variables in your Lasso model after excluding the originally identified top five predictors, we need to retrain the model and evaluate its feature importances. Here's the general process to follow:

1. Rebuild the Lasso model :- When we train the model using the updated dataset that excludes the five most important predictor variables from the original model.
2. Evaluate feature importance :- after training the model the absolute The absolute values of these coefficients can give we an idea of the importance of each predictor in the new model. Larger absolute coefficients indicate greater importance.
3. the importance of predictor variables can change when we modify the dataset or the model, so it's crucial to re-evaluate the feature importances in the context of your updated model and data

Q4- 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?

Ans:- robustness and generalization are essential qualities of a machine learning model. While they may impact training data accuracy, they ensure that the model can perform well in practical, real-world situations and generalize effectively to new, unseen data.

1. Accuracy on test data :
2. Bais variance
3. **Balancing Precision and Recall:** a more robust model may sacrifice some precision or recall to achieve better overall generalization. It might make fewer false predictions (higher precision) at the cost of missing some true positives (lower recall), or vice versa.
4. **Monitor Performance:** Continuously monitor the model's performance in production. Use feedback loops and retraining mechanisms to adapt to changing data patterns and maintain robustness over time.
5. **Evaluate on Diverse Data:** Test the model on diverse and representative data. Ensure that the testing data is similar to the real-world data that the model will encounter. This helps you assess how well the model generalizes to different scenarios.

Special note:- to be honest I m not much good in programming language but all things have been done with full on effort , and I m doing hard work to do best