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?

Ans:- **Optimal Alpha for Ridge and Lasso Regression:**

To find the optimal alpha values for ridge and lasso regression, can use cross-validation techniques such as Grid SearchCV or Randomized SearchCV. These techniques involve trying different alpha values and selecting the one that minimizes the mean squared error (MSE) or another chosen evaluation metric.

Impact of Doubling Alpha:

After determining the optimal alpha values, can double these values for both ridge and lasso regression. Then, retrain the models using the doubled alpha values. Compare the performance metrics (such as MSE or R-squared) of the updated models with those of the original models to assess the impact of doubling alpha.

Identifying Important Predictor Variables:

To identify the most important predictor variables, examine the coefficients of the predictor variables in both the original and updated models. For ridge regression, the coefficients will shrink but not necessarily become zero. For lasso regression, some coefficients may become exactly zero, indicating that the corresponding predictor variables are not contributing to the model. Plotting the coefficients before and after doubling alpha can help visualize the changes.

Visualization:

Visualize the coefficients of predictor variables using bar plots or heatmap plots for both the original and updated models. Additionally, plot the performance metrics (e.g., MSE) for different alpha values to visualize the impact of regularization strength on model performance.

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?

Ans :-Optimal Value of Lambda (Alpha):

Have determined the optimal value of alpha for both ridge and lasso regression using cross-validation techniques like GridSearchCV or RandomizedSearchCV.

Choice between Ridge and Lasso Regression:

If the optimal value of alpha for lasso regression results in some coefficients being exactly zero, indicating variable selection, and this aligns with the objective of simplifying the model and selecting important features, may choose lasso regression.

Conversely, if retaining all features and reducing multicollinearity is a priority, you may choose ridge regression.

Decision Criteria:

Consider the interpretability of the model: Lasso regression performs feature selection by shrinking some coefficients to zero, which may lead to a more interpretable model with fewer variables.

Consider the complexity of the model: Ridge regression tends to shrink all coefficients towards zero without necessarily setting any to zero. If the goal is to maintain all features while reducing overfitting, ridge regression might be preferred.

Consider the trade-off between bias and variance: Lasso regression tends to perform better in situations where there are a large number of irrelevant features, as it can effectively eliminate them. However, if there are strong multicollinearity issues, ridge regression might be more appropriate.

Final Choice:

Based on the analysis of your specific dataset and the objectives of modeling task, choose either ridge or lasso regression based on the trade-offs mentioned above and the specific requirements of your project.

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?

Ans **Identification of Important Predictor Variables:**

Have already identified the five most important predictor variables in the lasso model during the model evaluation phase.

Exclusion from Incoming Data:

Now, if these five important predictor variables are not available in the incoming data, you need to exclude them from the feature set before building the new model.

Identification of New Important Predictor Variables:

After excluding the five most important predictor variables, you need to retrain the model and identify the new set of important predictor variables using the lasso regression technique.

Evaluation and Comparison:

Compare the new set of important predictor variables with the original five to assess the impact of excluding them on model performance and interpretability.

Reporting:

Report the new set of important predictor variables and discuss their implications for the model and the analysis.

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?

Ans Cross-Validation: utilized k-fold cross-validation to assess the performance of your models, such as Ridge and Lasso regression.

Feature Selection and Regularization: employed Ridge and Lasso regression techniques to perform feature selection and regularization, which help in building simpler and more generalizable models.

Testing on Independent Data: split the data into training and validation sets and evaluated the model's performance on the validation set, which serves as an approximation of unseen data.

Outlier Detection and Handling: While not explicitly mentioned, it's assumed that addressed outliers during data preprocessing or model training to ensure robustness.

Model Evaluation Metrics: likely used evaluation metrics such as Mean Squared Error (MSE), R-squared, and Root Mean Squared Error (RMSE) to assess the model's performance and generalization ability.

Consistent Performance Across Multiple Datasets: may have evaluated the model's performance across different subsets of the data or through repeated cross-validation to ensure consistency and generalization.

Sensitivity Analysis: might have conducted sensitivity analysis by varying input parameters or data distributions to understand the model's stability and generalization across different scenarios.