L1 and L2 Regularization Assignment

Problem Statement:
This assignment aims to understand and implement L1 and L2 regularization techniques in regression models to mitigate overfitting and improve model generalization. By utilizing L1 (Lasso) and L2 (Ridge) regularization, students are expected to build regression models, compare their performance with non-regularized models, and interpret the effects of regularization parameters on model coefficients.

Guidelines:
1. Foundational Knowledge:
   - Understand the concepts of L1 (Lasso) and L2 (Ridge) regularization in regression.
   - Familiarize yourself with the mathematical formulations of L1 and L2 regularization.
   - Recognize the role of regularization in controlling model complexity and preventing overfitting.

2. Data Exploration:
   - Analyze the dataset's structure and characteristics using exploratory techniques like histograms, scatter plots, and correlation matrices.
   - Identify potential predictors and their relationships with the target variable.

3. Preprocessing and Feature Engineering:
   - Handle missing values and categorical variables appropriately.
   - Encode categorical variables if necessary.
   - Scale or normalize features to ensure regularization effectiveness.

4. Regression Model Construction:
   - Choose appropriate regression model for comparison.
   - Implement L1 and L2 regularization techniques using suitable libraries.
   - Tune regularization parameters based on cross-validation or validation set performance.

5. Model Evaluation:
   - Evaluate the performance of regularized regression models using metrics such as mean squared error, mean absolute error, and R-squared score.
   - Compare the performance of regularized models with non-regularized models to understand the impact of regularization on model performance.

6. Hyperparameter Tuning and Model Optimization:
   - Perform hyperparameter tuning for regularization strength using techniques like grid search or random search.
   - Validate the optimized models using cross-validation techniques to ensure generalization.

Step-by-Step Approach to Regularization Modeling:

1. Setup and Data Preparation:
   - Import necessary libraries: pandas, matplotlib, scikit-learn.
   - Load the dataset for regression modeling.
   - Preprocess the data, handle missing values, and encode categorical variables.

2. Regression Models:
   - Implement L1 (Lasso) and L2 (Ridge) regression models using scikit-learn.

3. Model Training and Evaluation:
   - Train both L1 and L2 models on the prepared training data.
   - Evaluate the performance of models on the testing set using relevant metrics.

4. Comparison and Analysis:
   - Compare the performance of the models.
   - Analyze the impact of regularization on model coefficients and overall performance.

5. Hyperparameter Tuning:
   - Perform hyperparameter tuning for regularization strength using cross-validation.
   - Select the optimal values for L1 and L2 regularization.

Links to Datasets for the Assignment:

- CO2 emission of cars dataset
[https://www.kaggle.com/datasets/midhundasl/co2-emission-of-cars-dataset/data]
- Paris Housing Price Prediction
[https://www.kaggle.com/datasets/mssmartypants/paris-housing-price-prediction/data]
- Car-Prices-Prediction-data
[https://www.kaggle.com/datasets/mrsimple07/car-prices-prediction-data/data]