Lab Assignment on Linear Regression

Made By : Arnav Bansal

Group : CSE-G1

SID : 23103046

Q1: Implement Linear regression, Ridge Regression, and LASSO regression on given dataset

after dividing it into a 70-30% ratio. Where 70% will be used to fit the regression line and 30%

will be used to test the model.

(a) Calculate MSE, and MAE of all three models and compare them.

(b) Print the coefficients of all three models after fitting.

Explanation :

Step 1: Dataset Division

* 70-30% Split:
  + 70% of the dataset will be used for training the model.
  + 30% of the dataset will be used to test the model.
  + You can use functions like train\_test\_split from scikit-learn to split the data.

Step 2: Implement Regression Models

1. Linear Regression:
   * This is a basic regression model that fits a linear relationship between the input variables and the output.
   * You will use Ordinary Least Squares (OLS) to minimize the difference between the predicted and actual values.
2. Ridge Regression:
   * Ridge is a type of linear regression but with L2 regularization, which helps avoid overfitting by penalizing large coefficients.
   * It introduces a penalty term proportional to the sum of squared coefficients.
3. LASSO Regression:
   * LASSO (Least Absolute Shrinkage and Selection Operator) is a type of linear regression with L1 regularization, which has the added effect of driving some coefficients to be exactly zero, thus performing feature selection.

Step 3: Fit the Models

* Fit all three models on the training set (70%) using the respective methods (fit() function).
* Each model will produce a set of coefficients after fitting the model, which describe the relationship between input features and the output.

Step 4: Calculate MSE and MAE

* After training, test the models on the remaining 30% of the dataset.
* Calculate the following metrics on the test data to evaluate the models:
  1. Mean Squared Error (MSE): This measures the average squared difference between the actual and predicted values. It is more sensitive to outliers.
  2. Mean Absolute Error (MAE): This measures the average absolute difference between the actual and predicted values. It is less sensitive to outliers compared to MSE.

These metrics will allow you to compare the performance of the models.

Step 5: Print Coefficients

* Once the models are trained, print the coefficients for each of the models.
* This gives insights into how the features impact the predictions.

CODE :

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression, Ridge, Lasso

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error

# Load your dataset (replace 'your\_dataset.csv' with the actual file path)

data = pd.read\_csv('friedman.csv')

# Features are Input1 to Input5, and the target is Output

X = data[['Input1', 'Input2', 'Input3', 'Input4', 'Input5']] # Features

y = data['Output'] # Target variable

# Split the dataset into training (70%) and testing (30%) sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Initialize the models

linear\_model = LinearRegression()

ridge\_model = Ridge(alpha=1.0) # Adjust alpha for Ridge regularization

lasso\_model = Lasso(alpha=1.0) # Adjust alpha for LASSO regularization

# Fit the models on the training data

linear\_model.fit(X\_train, y\_train)

ridge\_model.fit(X\_train, y\_train)

lasso\_model.fit(X\_train, y\_train)

# Make predictions on the testing data

y\_pred\_linear = linear\_model.predict(X\_test)

y\_pred\_ridge = ridge\_model.predict(X\_test)

y\_pred\_lasso = lasso\_model.predict(X\_test)

# Calculate MSE and MAE for Linear, Ridge, and LASSO regression

mse\_linear = mean\_squared\_error(y\_test, y\_pred\_linear)

mae\_linear = mean\_absolute\_error(y\_test, y\_pred\_linear)

mse\_ridge = mean\_squared\_error(y\_test, y\_pred\_ridge)

mae\_ridge = mean\_absolute\_error(y\_test, y\_pred\_ridge)

mse\_lasso = mean\_squared\_error(y\_test, y\_pred\_lasso)

mae\_lasso = mean\_absolute\_error(y\_test, y\_pred\_lasso)

# Print the results

print("Linear Regression: MSE =", mse\_linear, ", MAE =", mae\_linear)

print("Ridge Regression: MSE =", mse\_ridge, ", MAE =", mae\_ridge)

print("LASSO Regression: MSE =", mse\_lasso, ", MAE =", mae\_lasso)

# Print the coefficients of all models

print("Linear Regression Coefficients:", linear\_model.coef\_)

print("Ridge Regression Coefficients:", ridge\_model.coef\_)

print("LASSO Regression Coefficients:", lasso\_model.coef\_)

