CS 4375

ASSIGNMENT 1

Names of students in your group:

Saidarsh Tukkadi

Caleb Kim

Number of free late days used: 0

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

https://datasetsearch.research.google.com/

https://archive.ics.uci.edu/dataset/9/auto+mpg

https://www.ibm.com/topics/gradient-descent

Overview

This report summarizes results from utilizing gradient descent for linear regression on the Auto MPG dataset. Optimizing the model to minimize the mean squared error (MSE) and maximize the R-squared (R^2) value involved adjusting hyperparameters such as learning rate and number of iterations.

Dataset

The Auto MPG dataset, obtained from the UCI Machine Learning Repository, is the dataset utilized in this assignment. The miles per gallon (MPG) performance of the several car models is included, which was the target variable for this regression analysis.

Model and Hyperparameters

The SGDRegressor from scikit-learn was used to perform linear regression. The following hyperparameters were tested:

Learning Rates: 0.01, 0.001, 0.0001
Iterations: 1000, 5000, 10000

Results Log

Trial 1: Learning Rate: 0.01, Iterations: 1000, MSE: 12.174004892080694, R-squared: 0.7614838209260087
Trial 2: Learning Rate: 0.01, Iterations: 5000, MSE: 10.284404085852554, R-squared: 0.7985053572463904
Trial 3: Learning Rate: 0.01, Iterations: 10000, MSE: 9.902657409066995, R-squared: 0.8059846345695263
Trial 4: Learning Rate: 0.001, Iterations: 1000, MSE: 11.15594006708611, R-squared: 0.7814300041467251
Trial 5: Learning Rate: 0.001, Iterations: 5000, MSE: 11.036746786807187, R-squared: 0.7837652690029039
Trial 6: Learning Rate: 0.001, Iterations: 10000, MSE: 10.979406986142877, R-squared: 0.7848886848619046
Trial 7: Learning Rate: 0.0001, Iterations: 1000, MSE: 11.57498344151813, R-squared: 0.7732200005019251
Trial 8: Learning Rate: 0.0001, Iterations: 5000, MSE: 11.580004048545486, R-squared: 0.7731216355008089
Trial 9: Learning Rate: 0.0001, Iterations: 10000, MSE: 11.568919905357266, R-squared: 0.7733387988340752

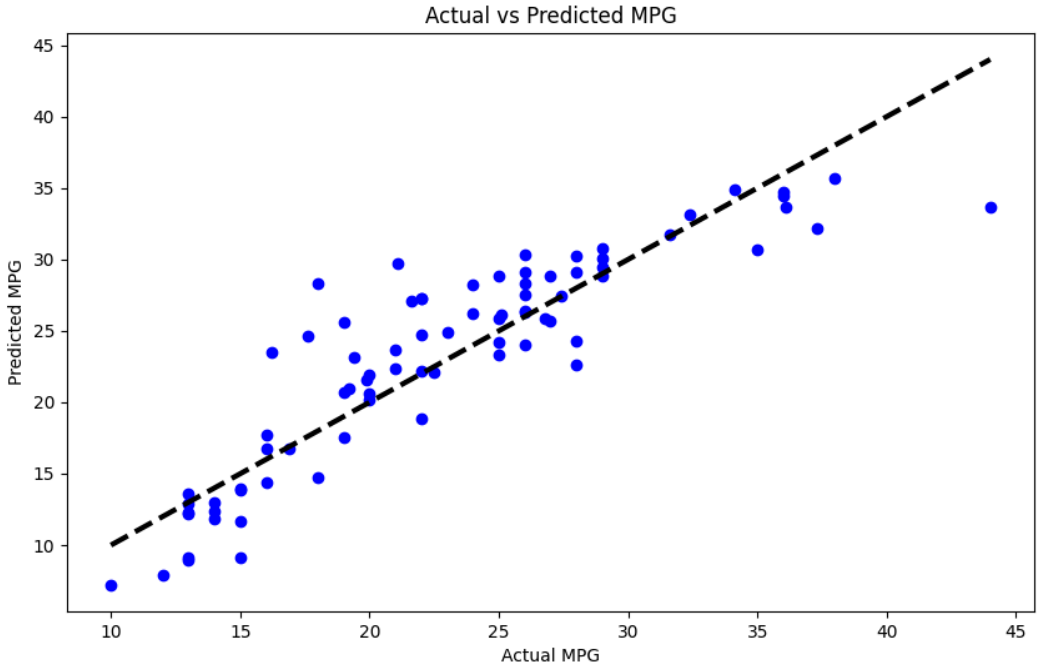
Analysis

Based on the results, there is a noticeable variation in the model's performance when the learning rate and iteration count are modified. Trial 3 displayed the best performance, with 10,000 iterations and a learning rate of 0.01 yielding the lowest MSE of 9.9027 and the highest R-squared of 0.8060. When compared to other trials, this indicates a better fit.

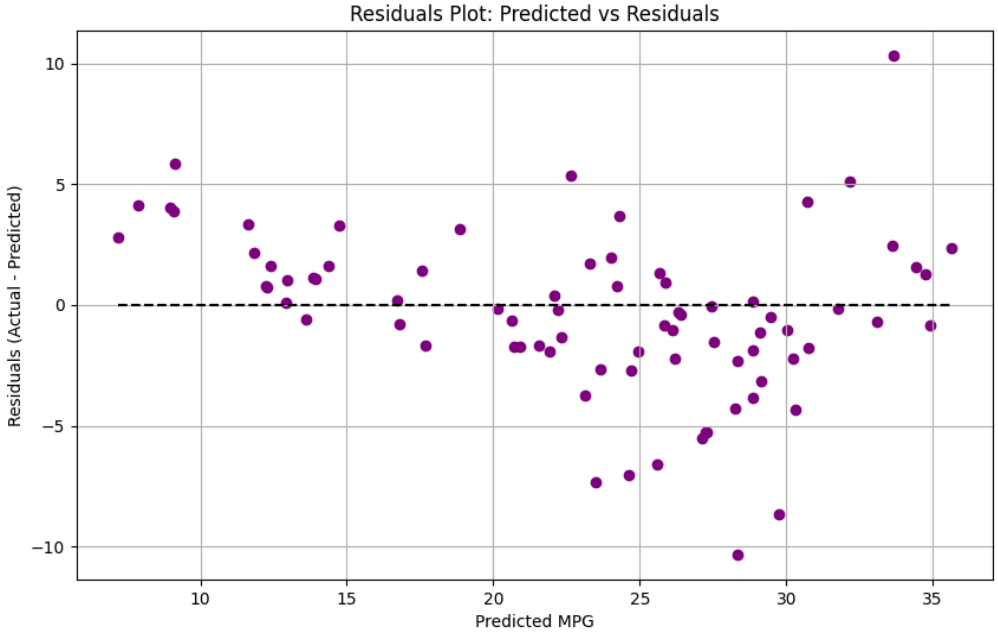
Are you satisfied that you have found the best solution?

Yes, the configuration from Trial 3 (Learning Rate: 0.01; Iterations: 10,000) offers the best balance between maximizing the explanatory power of the model and minimizing error. This setup ideally balances learning rate and iterations to achieve an effective model fit, as indicated by the relatively low MSE and high R-squared.

Plots



Actual vs. Predicted MPG: This plot shows how closely the predicted MPG values align with the actual values from the test set.



Residuals Plot: A residuals plot was generated to examine the distribution of residuals. Ideally, residuals should be randomly distributed around zero, indicating that the model predictions are unbiased.