

MPG Prediction Project - Explanation

1. Introduction

This project involves predicting the fuel efficiency (MPG - Miles Per Gallon) of cars based on features like horsepower and weight using linear regression. It uses data from the UCI Auto MPG dataset.

2. Steps Involved

- Load and clean the dataset (handle missing values, convert types).
- Visualize data with scatter plots to see relationships.
- Train a Linear Regression model using scikit-learn.
- Evaluate the model using MSE and RMSE.
- Predict MPG values based on user input.

3. Libraries Used

- pandas – for data manipulation
- matplotlib – for visualization
- sklearn – for machine learning (Linear Regression, metrics)

4. Linear Regression Explanation

Linear regression models the relationship between input features and a target variable. In this case, we predict MPG using horsepower and weight. The model learns the best-fit line through the data using the formula:

$$\text{MPG} = a * \text{horsepower} + b * \text{weight} + c$$
 (where a, b are coefficients and c is the bias/intercept).

5. Model Training

The model is trained using:

```
lr = LinearRegression()  
lr.fit(filtered_cars[['horsepower', 'weight']], filtered_cars['mpg'])
```

Here, the model learns the values of coefficients and bias using the data.

6. Model Evaluation

Why evaluate the model?

Helps us measure how well our predictions match real MPG values.

Metrics Used:

- Mean Squared Error (MSE): Measures the average squared difference between actual and predicted MPG values. Lower MSE means better accuracy.
- Root Mean Squared Error (RMSE): Square root of MSE. Helps us understand the average error in predictions.

7. Evaluation Result Example

- MSE = 17.84
- RMSE = 4.22

This means the average prediction error is about 4.22 MPG.

8. Predicting MPG from User Input

After training, the model can predict MPG for new values:

- User provides horsepower and weight.
- The model predicts the corresponding MPG using the learned equation.

9. Conclusion

This project demonstrates how linear regression can be applied to real-world datasets for prediction. By preprocessing data, visualizing relationships, training models, and evaluating performance, we can gain insights and make accurate predictions.