CYBR 486 - Lab #3: Linear Regression Model

Overview

This lab focuses on building, training, and evaluating a linear regression model using Python and scikit-learn. You will work with the Boston Housing dataset to predict housing prices. Key steps include preprocessing the dataset, splitting it into training and testing subsets, training the model, making predictions, and evaluating performance using various metrics.

Objectives

1. Load and preprocess the dataset.
2. Split the dataset into training and testing subsets.
3. Train a linear regression model.
4. Generate predictions using the trained model.
5. Evaluate the model using:
   * Mean Absolute Error (MAE)
   * Mean Squared Error (MSE)
   * Root Mean Squared Error (RMSE)
   * R² Score

Prerequisites

1. Python 3.x installed on your machine.
2. Required Python libraries:
   * scikit-learn
   * pandas
   * numpy

To install the required libraries, run the following command:

bash

Copy code

pip install scikit-learn pandas numpy

Dataset

* File: BostonHousing.csv
* The dataset contains information about various features of houses (e.g., number of rooms, property age) and the corresponding prices.

Ensure the dataset is placed in the same directory as the code file before running the script.

Usage Instructions

1. Clone the Repository  
   Clone this repository or download the code file and dataset to your local machine.
2. Run the Script  
   Execute the Python script (linear\_regression\_lab.py) using the command:

bash

Copy code

python linear\_regression\_lab.py

1. Outputs
   * Data types and null value checks are displayed in the console.
   * Training and test dataset shapes are printed.
   * Model evaluation metrics (MAE, MSE, RMSE, R²) are displayed in the console.
2. Save Results (Optional)  
   The evaluation metrics are saved to a CSV file named model\_evaluation\_results.csv in the same directory.

Code Structure

* Step 1: Import necessary libraries.
* Step 2: Load and preprocess the dataset (check for nulls, separate features and target).
* Step 3: Split the data into training and testing subsets.
* Step 4: Train the linear regression model and make predictions.
* Step 5: Evaluate the model using various metrics.

Evaluation Metrics

1. Mean Absolute Error (MAE): Average magnitude of errors.
2. Mean Squared Error (MSE): Average squared error.
3. Root Mean Squared Error (RMSE): Square root of MSE, indicating error magnitude in the same units as the target variable.
4. R² Score: Proportion of variance explained by the model (1.0 = perfect prediction)