Name: Vivek Mule
Roll No: 281072
Batch: A3

Practical 6

Practical Overview

This practical demonstrates the application of Linear Regression to predict month-wise temperature trends using historical temperature data. The goal is to forecast average monthly temperatures recorded at various locations in India. The project includes data preprocessing, model training, performance evaluation, and visualization to assess how well the regression model fits the data.

Objectives

- **Develop a Predictive Model:**
  Implement Linear Regression using a suitable library function to forecast month-wise temperatures.

- **Data Preparation:**

  - Load the temperature dataset using Pandas.

  - Address missing values and analyze the data distribution.

  - Split the dataset into training (e.g., 80%) and testing (e.g., 20%) sets.

- **Model Evaluation:**

  - Compute performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Square Score to evaluate the model's performance.

  - Visualize the model's predictions alongside the actual temperature values.

- **Visualization:**
  Plot the regression line along with the actual versus predicted temperature values using Matplotlib to assess the model fit.

Resources Utilized

- **Software:**

  - Visual Studio Code

- **Libraries:**

  - NumPy

  - Matplotlib

  - Scikit-learn

Theoretical Background

- **Linear Regression:**
  Linear Regression is a statistical technique for forecasting continuous outcomes. It models the relationship between one or more independent variables (X) and a dependent variable (Y). There are two main types:

  - *Simple Linear Regression:* Uses one independent variable.

  - *Multiple Linear Regression:* Involves multiple independent variables.

- **Applications:**
  It is widely used in forecasting and estimation tasks, such as predicting temperature, student performance, or sales figures, based on historical trends.

- **Limitations:**

  - Assumes a linear relationship between variables, which may not always be valid.

  - Sensitive to outliers that can skew the results.

  - Does not imply causation even when a strong correlation is present.

Methodology

1. **Data Preprocessing:**

   - **Loading and Inspection:**
     Load the temperature dataset using Pandas and inspect the data for any inconsistencies or missing values.

   - **Handling Missing Values:**
     Identify and address missing data appropriately to ensure a high-quality dataset.

   - **Visualization and Analysis:**
     Analyze the distribution of temperature values to understand underlying

trends.

- ○ **Data Splitting:**
  Divide the dataset into training and testing sets to enable effective model validation.

2. **Model Training:**

- ○ **Implementation:**
  Utilize Scikit-learn's Linear Regression function to build the model.

- ○ **Fitting the Model:**
  Train the model using the training dataset and predict month-wise temperatures on the testing set.

3. **Performance Evaluation:**

- ○ **Metrics Calculation:**
  Assess the model's performance using:

  - ■ **Mean Squared Error (MSE):** Measures the average squared differences between actual and predicted values.

  - ■ **Mean Absolute Error (MAE):** Measures the average absolute differences.

  - ■ **R-Square Score:** Indicates the proportion of variance in the dependent variable that is predictable from the independent variable(s).

4. **Model Visualization:**

- ○ **Graphical Representation:**
  Plot the actual versus predicted temperature values along with the regression line using Matplotlib to visually assess the model's fit.

Conclusion

The project successfully applied a Linear Regression model to predict month-wise temperatures using historical data. The model was evaluated through key performance metrics—MSE, MAE, and R-Square Score—which provided insights into its predictive capabilities. The visualizations further helped in understanding the regression line's alignment with the actual data. Future work could include exploring polynomial regression and further feature engineering to enhance model accuracy and robustness.