Assignment - 6

Name: Yashashree Nimbalkar
Roll No: 281040
Batch: A2
PRN: 22311423

Statement

In this assignment, we aim to:

a) Apply Linear Regression to predict month-wise temperatures in India using historical data.
b) Evaluate the regression model's performance using:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- R-Squared (R^2) Score
  c) Visualize the regression model using appropriate plots.

Objective

1. Understand regression techniques in machine learning.
2. Implement Linear Regression for predictive analysis.
3. Evaluate model performance using key regression metrics.
4. Visualize results for better interpretability.

Resources Used

- **Software:** VS Code

- **Libraries:** Pandas, NumPy, Matplotlib, Scikit-learn

Introduction to Regression Analysis

Regression analysis is a fundamental machine learning technique used to predict a continuous variable based on one or more input features. In this assignment, we applied Linear Regression to forecast monthly temperatures in India based on historical data.

Key Libraries Used:

1. **Pandas & NumPy:** Data loading and manipulation.
2. **Matplotlib:** Visualization of trends in temperature.
3. **Scikit-learn:** Model training, prediction, and evaluation.

Methodology

1. Data Collection and Preprocessing

- **Dataset Used:** Historical temperature data of India
- **Data Source:** Kaggle dataset (India temperature records)
- **Initial Steps:**
    - Loaded the dataset using Pandas.
    - Examined the first few rows to understand the structure.
    - Removed unnecessary columns (YEAR, ANNUAL) to focus on monthly temperature trends.

2. Splitting Data for Training and Testing

- **Defined Features (X) and Target Variable (y):**

- Features (X): Monthly temperature columns (Jan–Dec).
- Target Variable (y): Annual temperature (to observe trends).

- **Split Data:**
  - 80% Training Set, 20% Testing Set using train_test_split().

3. Model Selection and Training

- **Algorithm Used:** Linear Regression
  - Chosen for its simplicity and effectiveness in identifying temperature trends.
  - Trained the model using the fit() method.

4. Model Evaluation

- Calculated the following regression performance metrics:
  - **Mean Absolute Error (MAE):** Measures the average absolute difference between actual and predicted values.
  - **Mean Squared Error (MSE):** Measures the average squared difference, penalizing larger errors.
  - **R^2 Score:** Determines how well the independent variables explain the variance in the dependent variable.

5. Visualization of Regression Model

- Plotted actual vs. predicted values to assess model accuracy.
- Used scatter plots to visualize trends and residuals.

Advantages of Regression Analysis

1. Helps in identifying relationships between variables.

2. Useful for forecasting continuous values like temperature trends.

3. Provides interpretable coefficients for understanding feature importance.

Disadvantages

1. Assumes a linear relationship, which may not always be true.

2. Sensitive to outliers, which can distort predictions.

Conclusion

This assignment focused on implementing Linear Regression to predict monthly temperatures in India. The model was trained and evaluated using MSE, MAE, and R^2 Score to assess accuracy. By visualizing the regression line and predictions, we gained insights into temperature variations and model performance. These regression techniques are widely used in climate studies, sales forecasting, and financial analysis.