Name: Prachi Shedge

Roll No: 281050

Batch: A2

Assignment 6

Statement

In this assignment, we aim to:

a) Apply Linear Regression using a suitable library function to predict month-wise temperatures.

b) Assess the performance of the regression model using MSE, MAE, and R-Square metrics.

c) Visualize the simple regression model.

Objective

1. Understand how to analyze structured temperature data using Python.

2. Apply regression techniques to forecast temperature trends.

3. Evaluate the effectiveness of the predictive model using statistical metrics.

4. Develop visualization techniques for model performance analysis.

Resources Used

- Software: Jupyter Notebook / Visual Studio Code

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

Introduction to Regression Analysis

Regression analysis is a fundamental technique in machine learning used for predicting continuous values. In this assignment, we utilize Linear Regression to model temperature data for India and predict month-wise temperature trends.

Key Concepts:

- \*\*Linear Regression\*\*: Establishes a linear relationship between independent and dependent variables.

- \*\*Model Evaluation\*\*: Uses MAE, MSE, RMSE, and R-squared metrics to measure model accuracy.

- \*\*Data Visualization\*\*: Uses scatter plots and regression lines to analyze model predictions.

Basic Functions Used

1. `pd.read\_csv()` - Loads data from a CSV file into a Pandas DataFrame.

2. `drop(columns=[])` - Removes unnecessary columns from the dataset.

3. `train\_test\_split()` - Splits data into training and testing sets.

4. `LinearRegression()` - Implements a linear regression model.

5. `fit()` - Trains the model on training data.

6. `predict()` - Predicts target values based on input features.

7. `mean\_absolute\_error()` - Calculates MAE for model evaluation.

8. `mean\_squared\_error()` - Computes MSE and RMSE.

9. `r2\_score()` - Measures the proportion of variance explained by the model.

10. `scatter()` & `plot()` - Visualizes regression predictions.

Methodology

1. \*\*Data Collection and Exploration\*\*

- The dataset contains month-wise temperature values for India.

- The first step involves loading and inspecting the dataset.

2. \*\*Feature Selection & Preprocessing\*\*

- Exclude unnecessary columns (`YEAR` is not required for predictions).

- Define independent variables (`X`) and target variable (`y`).

3. \*\*Model Training\*\*

- Split data into training (80%) and testing (20%) sets.

- Train a Linear Regression model using the training data.

4. \*\*Model Evaluation\*\*

- Compute MAE, MSE, RMSE, and R-squared metrics.

- Compare actual vs. predicted values.

5. \*\*Visualization\*\*

- Create scatter plots to visualize regression performance.

- Plot an ideal fit line for reference.

Results & Observations

- Regression analysis provided meaningful predictions of month-wise temperatures.

- Evaluation metrics demonstrated model efficiency with a certain level of error.

- Visualization helped interpret the model's effectiveness.

Advantages of Regression Modeling

1. \*\*Predictability\*\*: Helps forecast temperature trends effectively.

2. \*\*Model Simplicity\*\*: Linear regression is easy to implement and interpret.

3. \*\*Data Insights\*\*: Provides a clear understanding of temperature variations.

Disadvantages

1. \*\*Assumptions\*\*: Linear regression assumes a linear relationship, which may not always hold.

2. \*\*Accuracy Limitations\*\*: Performance depends on data quality and feature selection.

Conclusion

This assignment demonstrated the application of Linear Regression in predicting month-wise temperatures. The analysis included data preprocessing, model training, evaluation, and visualization. The model provided valuable insights into temperature trends, highlighting the importance of regression in data science applications.