Name: Mayuresh Marade
Roll No: 281052
Batch: A2

Assignment 6

Statement

Q. Download the **temperatures dataset** from the following link:
Temperatures of India Dataset

The dataset contains **month-wise average temperatures across India** in Celsius.

Tasks:

a) Apply **Linear Regression** using a suitable library function to predict month-wise temperature
b) Assess the performance of regression models using **MSE**, **MAE**, and **R-Square** metrics
c) **Visualize** a simple regression model

Objective

This assignment helps in understanding:

1. The application of **Linear Regression**.
2. How to make **predictions** using linear models.
3. How to evaluate model performance with suitable metrics.

Resources Used

- **Software**: Python 3.x, Google Colab
- **Libraries**: NumPy, Matplotlib, Scikit-learn

Introduction to Linear Regression

1. **Linear Regression** is a statistical method for modeling relationships between variables.
2. It is mainly used to **predict continuous numeric outcomes**.
3. The method finds a **best-fit line** representing the relationship between an **independent variable (X)** and a **dependent variable (Y)**.

Types of Linear Regression

- **Simple Linear Regression**: Uses a single feature to predict the target variable.
- **Multiple Linear Regression**: Uses multiple features for prediction.

Applications of Simple Linear Regression

1. **Predicting student grades** based on study hours.
2. **Estimating agricultural yield** based on rainfall.

3. **Forecasting salary** based on years of experience.

Limitations of Simple Linear Regression

1. Assumes a **linear** relationship between features.
2. **Sensitive to outliers**, which can distort predictions.
3. **Does not imply causation**, only correlation.

Methodology

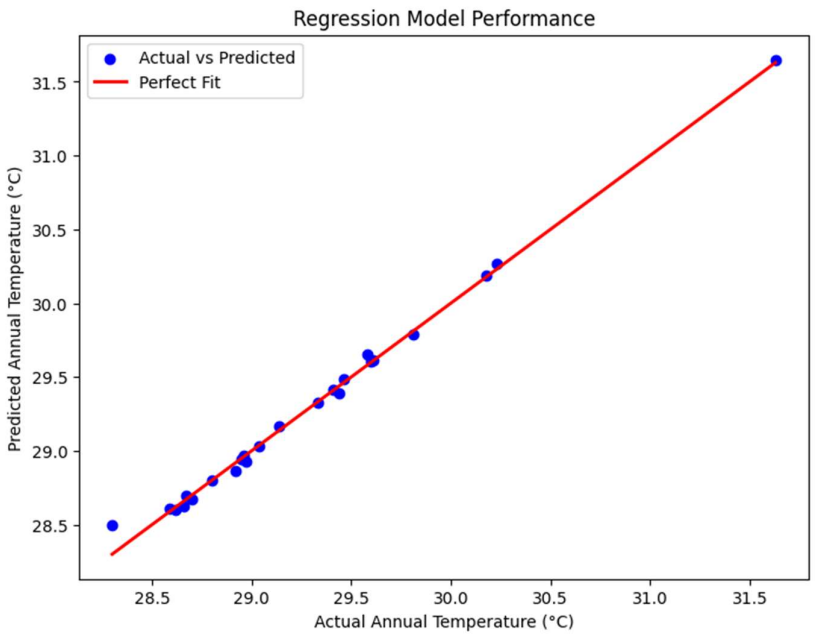
1. **Import Libraries**: Load NumPy, Pandas, Matplotlib, and Scikit-learn.
2. **Data Collection**: Load the temperature dataset.
3. **EDA & Preprocessing**:
    - Check for missing values
    - Visualize monthly trends
4. **Data Splitting**:
    - Use train_test_split to divide data into training and testing sets
5. **Model Training**:
    - Use LinearRegression() to train on the data
6. **Predictions**:
    - Predict temperatures for the test set
7. **Model Evaluation**:
    - Calculate **Mean Squared Error (MSE)**
    - Calculate **Mean Absolute Error (MAE)**
    - Compute **R² Score**

Results

```
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R-Square: {r2}')
```

```
MSE: 0.002440835094691285
MAE: 0.029590158623251355
R-Square: 0.9949515631120867
```



Conclusion

A **simple linear regression model** successfully captures the **linear relationship** between month and average temperature in India. While simple and interpretable, such models serve as a foundation for more complex regression-based forecasting in domains like **weather analysis**, **agriculture**, and **finance**.