Name: Manisha Apchunde
PRN: 22420244
Roll no: 281076
Batch:A3

Assignment 06

Aim:

Download temperature data from below link.
https://www.kaggle.com/venky73/temperaturesof-india?select=temperatures.csv
This data consists of temperatures of INDIA averaging the temperatures of all places month wise. Temperatures values are recorded in CELSIUS

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

Objective

To apply Linear Regression on the temperature dataset of India to predict month-wise temperatures and evaluate the model using performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Square. Additionally, the regression model will be visualized.

Libraries and Packages Used

pandas – for reading and handling the dataset numpy – for

numerical operations matplotlib and seaborn – for data

visualization sklearn.linear_model – for Linear Regression

model sklearn.metrics – to evaluate model performance using

MSE, MAE, and R-Square

Theory

What is Linear Regression?

Linear Regression is a supervised machine learning algorithm used for predicting a continuous dependent variable based on one or more independent variables. It attempts to find the best-fitting straight line
(regression line) through the data points, minimizing the difference between actual and predicted values.

Types of Linear Regression

Simple Linear Regression – involves one independent variable

Multiple Linear Regression – involves two or more independent variables

Polynomial Regression – a form of linear regression where the relationship is modeled as an nth degree polynomial

Applications of Simple Linear Regression :

Predicting sales based on advertising spend

Estimating house prices based on area

Predicting student scores based on study hours

Forecasting temperatures based on time or season

Limitations of Simple Linear Regression :

Assumes a linear relationship between variables

Sensitive to outliers

Not ideal for complex relationships or datasets with multicollinearity

Can underperform if important independent variables are omitted

Dataset

The dataset used was downloaded from Kaggle:

https://www.kaggle.com/venky73/temperaturesof-india

It contains the average temperatures of India recorded month-wise in Celsius.

Working of Linear Regression :

Data Collection :

The CSV file was loaded using pandas. It contains columns for years and temperature values.

Data Preprocessing :

Extracted relevant columns (e.g., year/month as independent variable and temperature as dependent variable)

Checked for and handled any missing or invalid values

Converted categorical month names to numerical values (if necessary)

Model Building :

Used LinearRegression() from sklearn.linear_model

Fitted the model on training data (months or years vs temperature)

Predicted temperatures for the months

Model Evaluation :

Used the following metrics from sklearn.metrics to assess model performance:

Mean Squared Error (MSE) – average of squared prediction errors

Mean Absolute Error (MAE) – average of absolute prediction errors

R-Square (R^2) – explains how well the regression line approximates real data points

Visualization :

Plotted the actual temperature data and the regression line using matplotlib and seaborn

Visual comparison helped verify the quality of predictions

Conclusion:

In this assignment, we applied simple linear regression on a real-world dataset of India's average monthly temperatures. The model was trained, tested, and evaluated using MSE, MAE, and R-Square. Visualization confirmed the trend and fit of the regression line. Linear regression proved effective in identifying and predicting trends in temperature data, although the model's simplicity may limit accuracy for highly nonlinear datasets.