** Bansilal Ramnath Agarwal Charitable Trust’s**

**Vishwakarma Institute of Information Technology, Pune-48**

**(An Autonomous Institute affiliated to Savitribai Phule Pune University)**

**Department of Computer Science and Engineering (Artificial Intelligence)**

**LAB SUBMISSION**

**Data Science and Machine Learning**

**CAUA22201**

*Submitted by:*

*Avdhoot Nilesh Pimparkar,*

*PRN: 22210178,*

*Roll Number: 281054*

*Second Year*

*Semester II Academic Year 2023-24*

**Assignment: 6**

Aim: To perform Regression technique on a dataset and

a) Apply Linear Regression using a 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 a simple regression model.

Theory:

Linear Regression: Linear regression is a fundamental technique used in predictive modelling to establish a relationship between a dependent variable (target) and one or more independent variables (features). In the context of temperature prediction from the provided dataset, linear regression can be applied to understand how changes in independent variables (such as month) affect the dependent variable (temperature).

To apply linear regression, a suitable library function can be used, such as scikit-learn in Python. The dataset can be pre-processed to extract the necessary features (e.g., month) and target variable (temperature). Then, the linear regression model can be trained on the training data using the fit() function. Once trained, the model can be used to predict the temperature for each month.

The performance of the regression model can be assessed using several metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Squared (R^2). MSE measures the average squared difference between the predicted and actual values, while MAE measures the average absolute difference. R^2 indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

Visualizing a simple regression model involves plotting the actual values against the predicted values. This can be done using scatter plots, where each point represents a data point, and the X-axis represents the actual values while the Y-axis represents the predicted values. A perfect regression model would result in all points lying on a diagonal line with a slope of 1. Visual inspection of the plot can provide insights into how well the model fits the data and whether there are any systematic errors or patterns in the predictions.

In summary, linear regression is a powerful technique for predicting temperature based on historical data. By assessing the performance of the regression model using appropriate metrics and visualizing the results, we can gain valuable insights into the accuracy and reliability of the predictions.

Here is the dataset that we have used:

“<https://www.kaggle.com/datasets/venky73/temperatures-of-india>”

Advantages:

1. Linear regression is a relatively simple algorithm, making it easy to understand and implement. The coefficients of the linear regression model can be interpreted as the change in the dependent variable for a one-unit change in the independent variable, providing insights into the relationships between variables.
2. Linear regression is computationally efficient and can handle large datasets effectively. It can be trained quickly on large datasets, making it suitable for real-time applications.
3. Linear regression is relatively robust to outliers compared to other machine learning algorithms. Outliers may have a smaller impact on the overall model performance.
4. Linear regression often serves as a good baseline model for comparison with more complex machine learning algorithms.
5. Linear regression is a well-established algorithm with a rich history and is widely available in various machine learning libraries and software packages.

Limitations:

1. Linear regression assumes a linear relationship between the dependent and independent variables. If the relationship is not linear, the model may not perform well.
2. Linear regression is sensitive to multicollinearity, which occurs when there is a high correlation between independent variables. Multicollinearity can inflate the variance of the coefficients and lead to unstable model predictions.
3. Linear regression assumes that the features are already in a suitable form for the model. Feature engineering may be required to transform features into a format that can be effectively used by the model.
4. Linear regression is susceptible to both overfitting and underfitting. Overfitting occurs when the model learns the training data too well and fails to generalize to unseen data. Underfitting occurs when the model is too simple to capture the underlying relationships in the data.
5. Linear regression provides limited explanatory power for complex relationships between variables. More advanced machine learning techniques may be necessary for deeper insights.

Applications:

Linear regression is used in many different fields, including finance, economics, and psychology, to understand and predict the behaviour of a particular variable. For example, in finance, linear regression might be used to understand the relationship between a company’s stock price and its earnings or to predict the future value of a currency based on its past performance.



Fig. Linear Regression

Results:

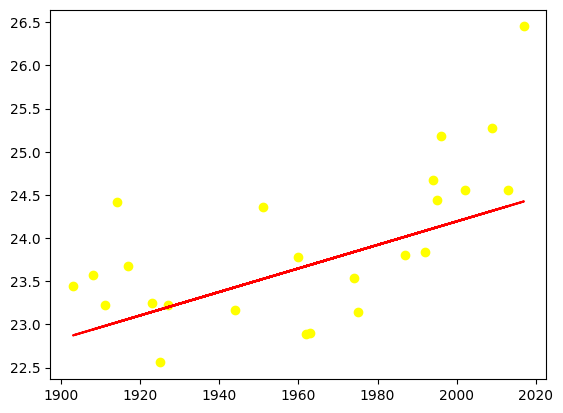
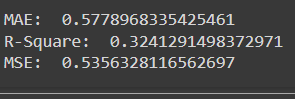
 

Fig. Linear Regression Fig. MAE, MSE & R-Square

Conclusion:

In this assignment, we were able to learn about Linear Regression and various errors like Mean Squared Error (MSE) and Mean Absolute Error (MAE) and R-Squared.