

Name: Dev Bhangale
Roll no: 281059
Batch: A3

Assignment – 6 **Regression technique**

Problem Statement:

Assignment on Regression technique.

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.

Libraries and packages used: NumPy, Matplotlib, scikit-learn

Theory:

Linear Regression

It's a statistical technique for forecasting analysis. Predictions are made using linear regression for continuous, real, or numerical variables like sales, earnings, age, and product price, among others.

The term "linear regression" refers to a procedure that displays a linear relationship between one or more independent (x) variables and a dependent (y) variable.

Given that linear regression displays a linear relationship, it can be used to determine how the value of the independent variable affects the value of the dependent variable.

Linear Regression Types:

Basic Linear Regression:

A linear regression procedure is referred to as simple linear regression if it uses one independent variable to predict the value of a number of dependent variables.

Multiple Linear Regression: This type of linear regression method is employed when multiple independent variables are combined to predict the value of a numerical dependent variable.

Applications of Simple Linear Regression:

1. Student grades determined by the number of hours studied (ideally): In this case, exam scores are dependent on the number of hours studied, but the number of hours studied is independent.

2. Estimating agricultural yields using rainfall data: The measure of precipitation is an independent variable, and yield is a dependent variable.
3. Estimating an individual's salary based on years of experience: Experience is now the independent variable, and salary is the dependent variable.

Limitations of Simple Linear Regression:

1. Assumes linearity: Regression models, particularly linear regression, assume a straight-line relationship between the independent and dependent variables. If the underlying relationship is more complex (curved, exponential, etc.), the model may not accurately capture the true association.
2. Sensitive to outliers: Outliers (data points significantly different from the majority) can disproportionately influence the regression line, leading to misleading results.
3. Doesn't establish causation: Even if a strong correlation is found between variables, regression models cannot determine causality. There might be a third, unseen variable influencing both the independent and dependent variables, creating a false association.

Working:

Linear regression is a fundamental technique in statistics and machine learning used to model the relationship between a dependent variable and one or more independent variables. The goal is to create a linear model that predicts the dependent variable based on the independent variables.

Here's how a linear regression model works for prediction:

- Importing necessary libraries and modules: They provide pre-built functionalities and extend your program's capabilities
- Data Collection: Collect data on the variables of interest. For example, in a simple linear regression, you would have one independent variable (for eg. Year here) and one dependent variable (for eg. Temperature here).
- Data Preprocessing and EDA: This step involves cleaning the data and analysing it intricately.
- Splitting the Data: Split the dataset into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance.
- Model Training: Use the training data to fit a linear regression model. The model tries

to find the best-fitting linear relationship between the independent and dependent variables. In simple linear regression, this relationship is represented by a line ($y = mx + b$), where m is the slope and b is the intercept.

- **Making Predictions:** Once the model is trained, use it to make predictions on the testing data. The model calculates the predicted values of the dependent variable based on the values of the independent variable(s).
- **Evaluating the Model:** Evaluate the model's performance using metrics such as mean squared error (MSE) or R-squared. These metrics measure how well the model's predictions match the actual values in the testing data.

Conclusion:

In summary, a simple linear regression model uses a straight line to determine the association between one independent variable and one dependent variable.