# Assignment No: - 6

## Regression

**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.

**Objective:** This assignment will assist us in understanding the applications of linear regression and how predictions can be made using it.

**S/W Packages and H/W apparatus used:** OS: Windows, Kernel: Python 3, Tools: Google Colab

**Libraries and packages used:** NumPy, Pandas, Matplotlib, Scikit-Learn

**Theory:**

**Linear Regression:** It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

Types of Linear Regression: -

- **Simple Linear Regression:** If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- **Multiple Linear regression:** If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression. Assumptions of Linear Regression To conduct a simple linear regression, one has to make certain assumptions about the data. This is because it is a parametric test.

The assumptions used while performing a simple linear regression are as follows:

- **Homogeneity of variance (homoscedasticity) -** One of the main predictions in a simple linear regression method is that the size of the error stays constant. This simply means that in the value of the independent variable, the error size never changes significantly.

- **Independence of observations -** All the relationships between the observations are transparent, which means that nothing is hidden, and only valid sampling methods are used during the collection of data.

- **Normality -** There is a normal rate of flow in the data. These three are the assumptions of regression methods.

  However, there is one additional assumption that has to be taken into consideration while specifically conducting a linear regression.

- **The line is always a straight line -** There is no curve or grouping factor during the conduction of a linear regression. There is a linear relationship between the variables (dependent variable and independent variable). If the data fails the assumptions of homoscedasticity or normality, a nonparametric test might be used. (For example, the Spearman rank test)

**Advantages and Disadvantages & Limitation/Example:**

**Advantages:**

1. **Quantifying Relationships:** Regression analysis allows us to quantify the relationship between one or more predictor variables and a response variable. This helps in

understanding how changes in the predictor variables affect the response variable, providing valuable insights into the underlying processes or phenomena being studied.

2. **Prediction:** Regression models can be used for prediction by estimating the response variable's value based on the values of predictor variables. This predictive capability is valuable in various fields such as finance, healthcare, and marketing for forecasting future trends or outcomes.

3. **Model Interpretability:** Linear regression models, in particular, offer straightforward interpretation of coefficients, indicating the magnitude and direction of the effect of each predictor variable on the response variable. This interpretability facilitates understanding and communication of the model's findings to stakeholders.

4. **Model Evaluation:** Regression models can be evaluated using various metrics such as R-squared, adjusted R-squared, and root mean squared error (RMSE) to assess their goodness of fit and predictive accuracy. This allows researchers to objectively evaluate the performance of the model and compare different models.

5. **Hypothesis Testing:** Regression analysis enables hypothesis testing to determine whether the relationships between predictor variables and the response variable are statistically significant. This helps in validating the assumptions underlying the model and assessing the reliability of the findings.

**Applications:**

1. **Marks scored by students based on number of hours studied (ideally) -** Here marks scored in exams are dependent and the number of hours studied is independent.

2. **Predicting crop yields based on the amount of rainfall -** Yield is a dependent variable while the measure of precipitation is an independent variable.

3. **Predicting the Salary of a person based on years of experience -** Therefore, Experience becomes the independent variable while Salary turns into the dependent variable.

**Limitations:**

Indeed, even the best information doesn't recount a total story. Regression investigation is ordinarily utilized in examinations to establish that a relationship exists between variables. However, correlation isn't equivalent to causation: a connection between two variables doesn't mean one causes the other to occur. Indeed, even a line in a simple linear regression

that fits the information focuses well may not ensure a circumstances and logical results relationship.

Utilizing a linear regression model will permit you to find whether a connection between variables exists by any means. To see precisely what that relationship is and whether one variable causes another, you will require extra examination and statistical analysis. Conclusion Simple linear regression is a regression model that figures out the relationship between one independent variable and one dependent variable using a straight line.

**Working / Algorithm:**

- 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.

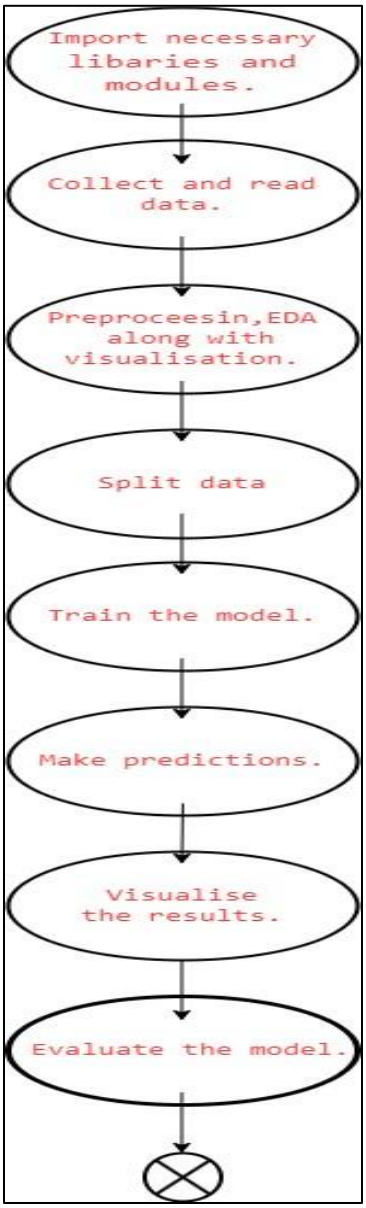
- Splitting the Data: Split the dataset into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance.

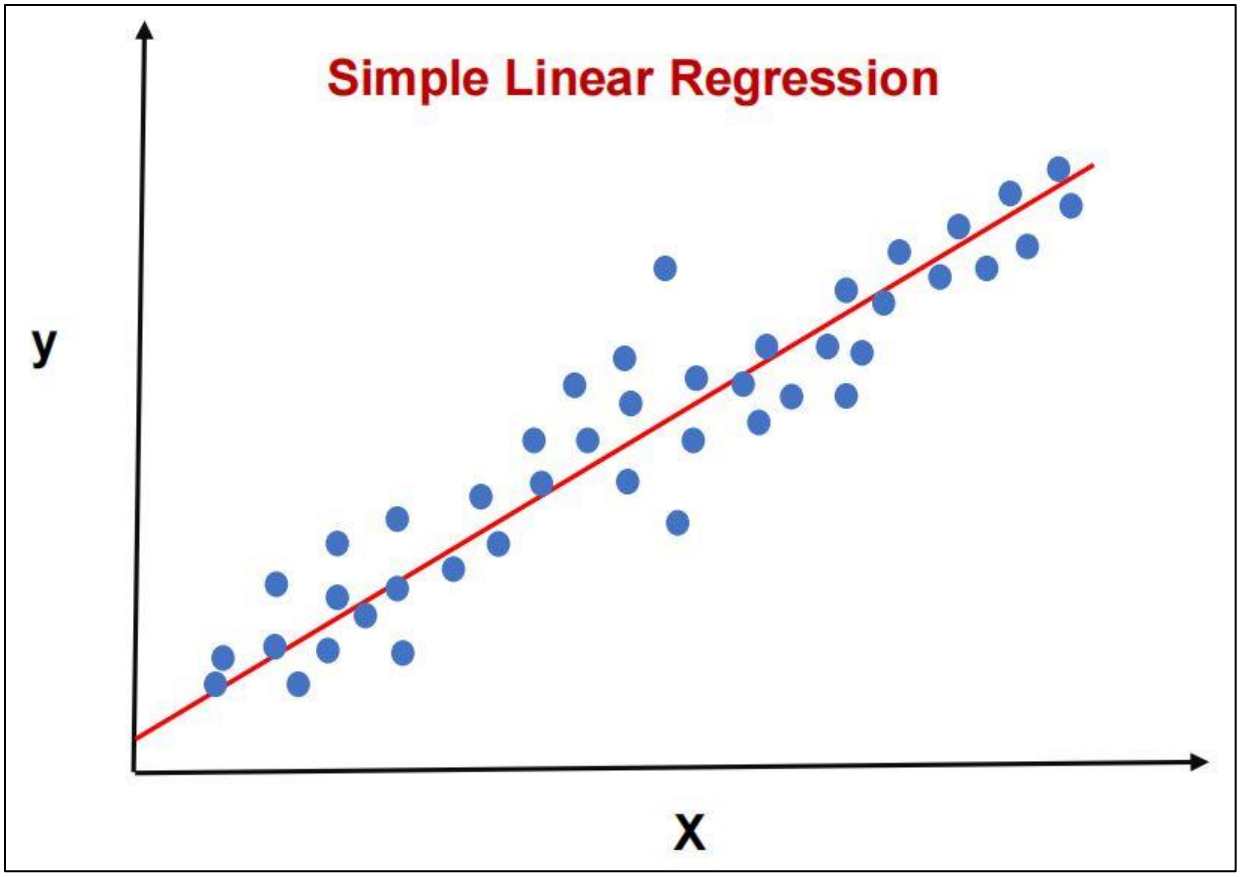
- Model Training: Use the training data to fit a linear regression model. The model tries to find the best-fitting linear relationship between the independent and dependent variables. In simple linear regression, this relationship is represented by a line ($y = mx + b$), where m is the slope and b is 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.

**Diagram:**


Import necessary libaries and modules.
Collect and read data.
Preproceesin,EDA along with visualisation.
Split data
Train the model.
Make predictions.
Visualise the results.
Evaluate the model.




**Conclusion:**

In summary, the application of linear regression on the temperature dataset from India facilitated the prediction of month-wise temperatures. Evaluation of regression model performance through metrics such as MSE, MAE, and R-Square provided insights into the accuracy and effectiveness of the predictions. Additionally, visualization of the regression model enhanced the understanding of the relationship between independent and dependent variables, aiding in interpretation and decision-making processes.

Simple linear regression is a regression model that figures out the relationship between one independent variable and one dependent variable using a straight line.