

## **Assignment: 1**

**Aim:** Assignment on Regression Technique.

This data consists of temperatures of INDIA over aging 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. Asses the performance of regression models using MSE,MAE and R-square metric's.

C. Visualize simple regression model

**Software used** – PyCharm Community Edition 2022.2.1

**Dataset:** Temperature (Temperature.csv)

**Link:** <https://www.kaggle.com/venky73/temperatures-of-india?Select=temperature.csv>

### **Theory:**

1:Regression: Regression is a method to determine the statistical relation ship between a dependent variable and one or more independent variables.

The change independent variable is associated with the change in Independent variables.

It's broadly classified into two types:

➤ **Linear Regression**

➤ **Logistic Regression**

a. **Linear Regression**- Linear regression is one of the easiest and

most popular machine learning algorithm.

It is a statistical method that is used for predictive analysis.

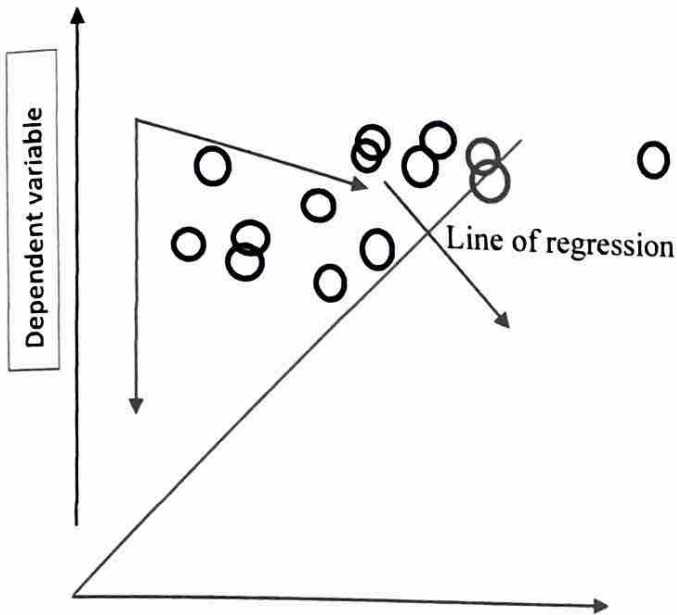
Linear regression makes predictions for continuous /real or numeric variables such are sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x)

Variables, hence called as linear regression.

Since linear regression shows the Linear relationships, which means it finds how the value of the independent variables.

Consider the below image:



Mathematically we can represent a linear regression as:

- Where,
- = Dependent variable (Target variable)
  - = Independent variable (Predictor variable)
  - = Intercept of the line (Gives an additional degrees of a freedom)
  - = Linear regression coefficient (scale factors to each input value)
  - = Random error.

The cost function for x and y variables are training datasets for linear regression model representation.

1. **Cost function:** A cost function establishes the relationship between events of a set of values to represent loss in gaining something.
2. **It is classified as:**
  - a. Mean Error (ME)
  - b. Mean Square Error (MSE)
  - c. Mean Absolute Error (MAE)
  - d. Root Mean Square Error (RMSE)
  - e. R-squared.

**a. Mean Error (ME):**

Mean error is the simplest of all cost functions.

It is a simply mean (average) of the difference between the actual value and the predicted value.

**b. Mean Square Error (MSE):**

It represents the average of the squared difference between the original and predicted values in the dataset.

It measures the variance of the residuals.

**Formula=**  $MSE = \frac{1}{N} \sum (y_i - \bar{Y})^2$

Where,

$\bar{Y}$  = Predicted value of y

$\bar{Y}$  = Mean value of y

**c. Mean Absolute error (MAE) :**

It represents the average of the absolute difference between the actual and predicted values in the dataset.

It measures the average of the residuals in the dataset.

**Formula:**  $MAE = \frac{1}{N}$

**d. Root Mean Square Error (RMSE) :**

It is the square root of Mean Square Error.

It measures the standard deviation of residuals.

**Formula:**

$$RMSE = \sqrt{\frac{1}{N} \sum (y_i - \bar{Y})^2}$$

**e. R-squared :**

It represents the proportion of the variance in the dependent variable which is explained by the linear regression model.

It is a scale free score i.e. irrespective of the values being small will be less than one

$$R^2 = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2} \cdot \frac{N \sum y^2 - (\sum y)^2}{N \sum y^2 - (\sum y)^2}$$

And for the adjusted R squared:

It is modified version of R square.

It is adjusted for the number of independent variables in the Model and it will always be less than or equal to  $R^2$ .

**Applications:**

**a. Simple linear regression:**

Marks scored by students based on the number of an hours studied (ideally):

Here marks are scored in exams are independent and the number of hours independent.

Predicting crop yields based on the amount of rainfall yield is a dependent variable while the measure of precipitation is an independent variable.

Predicting the salary of a person based on years of experience.

Therefore, Experience becomes the independent while salary turns into the dependent variable.



**b. Regression:**

1. Trend line
2. Epidemiology
3. Finance
4. Economics
5. Machine learning
6. Environmental Science

**Advantages:**

1. Predictive analytics
2. Operation efficiency
3. Supporting decisions
4. Correcting errors
5. New Insights

**Disadvantages:**

1. It involves very lengthy and complicated procedure of calculations and analysis.
2. It cannot be use in case of qualitative phenomenon  
Honesty crime etc.
3. It is assume that the cause and affect relationship between the variables remains unchanged.

**Examples:**

1. Simple linear regression
2. Linear regression
3. Logistic regression
4. Multiple linear regression

**Libraries use for this practical are:**

1. Pandas As pd
2. NumPy as np
3. Matplotlib pilots as plt
4. Sea born as seabornInstance
5. Sklearn

**Conclusion:**

Regression model is the relationship between are independents variable and one dependent variable using a straight line to figure out it, concluded it.