

ASSIGNMENT NO. 6

Title: Assignment on Regression technique.

Download temperature data from the link below.

<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 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.

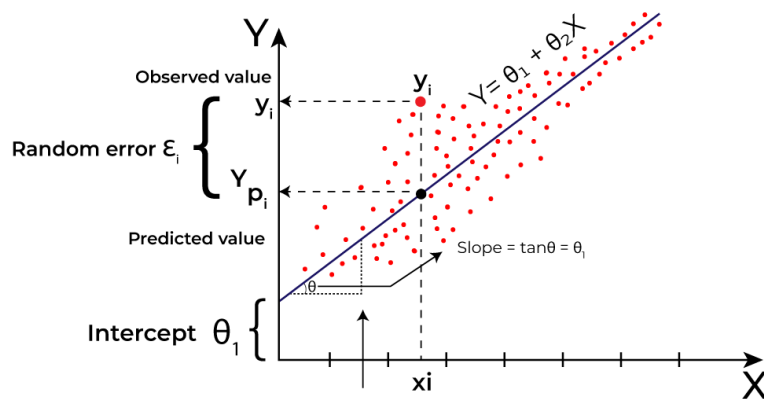
S/W Packages and H/W apparatus used: Linux OS: Ubuntu/Windows ,
Jupyter notebook.

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 (x) 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)

Applications of Simple Linear Regression

- 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 of Simple Linear Regression

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. By applying linear regression to historical temperature data, we can predict month-wise temperatures in India. Evaluation metrics such as MSE, MAE, and R-Squared provide insights into the performance of the regression model. Additionally, visualizing the regression model enhances our understanding of the relationship between variables and the accuracy of predictions.