ASSIGNMENT 2

Title : Regression technique.

Problem statement :
The data consists of temperatures of INDIA averaging the temperatures of all places monthwise. Temperatures values are recorded in CELSIUS.

a. Apply Linear Regression using suitable library fun^n. and predict the monthwise temperature.
b. Assess the performance of regression model using MSE, MAE and R-square metrics.
c. Visualize simple regression model.

Objective :
This assignment will help students to realize how linear regression can be used and predictions using same can be performed.

Theory :

- Definition of Linear Regression :
In layman terms, we can define linear regression as it is used for learning linear relationship between target and one or more forecasters, and it is probably one of the most popular and well inferential algorithms in statistics. Linear regression endeavours to demonstrate the connection between two variables by fitting a linear equation

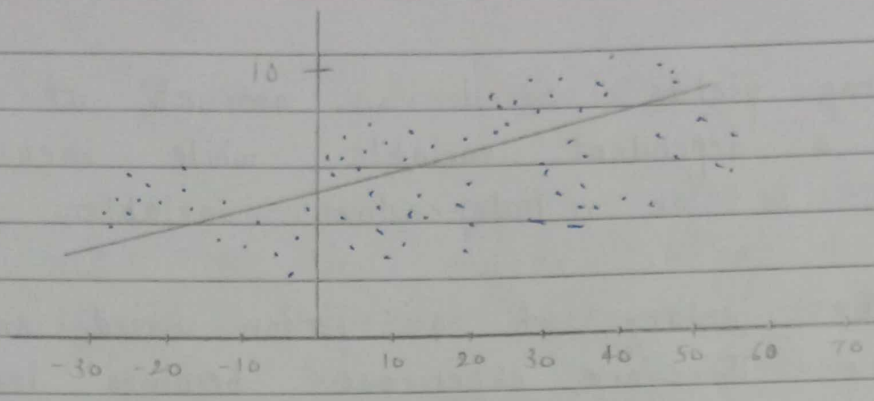
to observed information one ~~information~~ variable is viewed as an explanatory variable, and other is viewed as dependent variable.

* Types of Linear Regression :

1. Multiple Linear Regression -
In this type of linear regression, we always attempt to discover the relationship between two or more independant variables or inputs and the corresponding dependant variable or output and the independant variables can be either continuous or categorical.

2. Simple Linear Regression -
In simple linear regression, we aim to reveal the relationship between a single independant variable or you can say input and a corresponding dependant variable or output. We can discuss this in a simple line as $y = \beta_0 + \beta_1 x + \varepsilon$.

Here Y speaks to the output or dependant variable, β_0 and β_1 are 2 obscure constants that speak to intercept and coefficient that is slope separately, and error term is ε (Epsilon). We can also discuss this in form of a graph and here is a simple linear regression model graph.



* Assumptions of Linear Regression :

1. Homogeneity of variance - One of the main predictions in a simple linear regression method is that size of error stay constant. This simply means that in the value of independant variable, the error size never changes significantly.

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

3. Normality - There is a normal rate of flow in data.

* Applications of simple Linear Regression :

1. Marks scored by students based on number of hours studied (ideally) - Here marks scored in exams are dependant and number of hours studied is independant.

2. Predicting crop yields based on amount of rainfall - Yield is a dependant variable while measure of precipitation is an independant variable.

3. Predicting the salary of a person based on years of experience - Therefore, experience becomes independant while salary turns into dependant variable.

* Limitations of Simple Linear Regression :

Indeed, even best information doesn't recount a total story. Regression investigation is ordinarily utilized in examination to set up 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 on well may not ensure a circumstances and logical results relationship.

* Evaluation metrics for linear regression model :

Evaluation metrics are a measure of how good a model performs and how well it approximates relationship.

1. Means Squared Error (MSE):

The most common metric for regression tasks is MSE. It has convex shape. It is average of squared difference between predicted and actual value. Since it is differentiable and has a convex shape, it's easier to optimize.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

2. Mean Absolute error (MAE):

It is simply average of absolute difference between target value and value predicted by model. Not preferred in cases where outliers are prominent.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$

3. Root Mean squared error (RMSE):

RMSE is sq. root of average of squared difference of predicted and actual value. Residual are measure of how distant points are from regression lines. Thus RMSE measures scatter of these residual.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$

4. R squared error (R^2) :

R^2 score is a metric that tells performance of your model, not loss in an absolute sense, how many wells did your model perform. R^2 calculates how much regression line is better than mean line. R^2 squared is also known as coefficient of determination or sometimes also known as goodness fit.

Conclusion :

Simple Linear Regression is a regression model that figures relationship between one independant variable and one dependant variable using straight line.