Assignment No:5

Name- Kudharan Sumit Dattatraya
Class- TE
Div - 4
Subject - DSBDAL

Problem statement-

1. Implement Logistic Regression using python/R to perform classification on Social_Network_Ads.csv dataset.
2. Compute Confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, precision, Recall on the given dataset.

Theory-

1) Explain regression.

→ Regression is defined as a method of estimating the value of one variable when that of the other is known & the variables are correlated.

Regression analysis is used to predict or estimate one variable in terms of the other variables.

It is useful in statistical estimation of demand curves, supply curves, production function, cost function, etc.

Types of Regression

1. Simple Regression & Multiple Regression
2. Linear Regression & Non linear Regression.

2) Explain Logistic Regression.

→ Logistic Regression is supervised learning classification algorithm used to predict the probability of an output variable. The nature of dependent variable is such that there would be only two possible classes.

Logistic Regression classification -

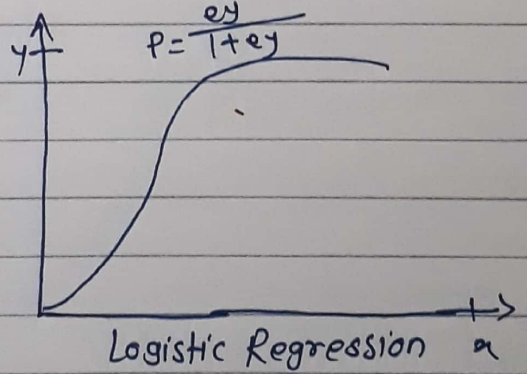
In a classification problem output or target variable y, can take any discrete values for given set of features or inputs X. L.R. is a regression model.

Steps of L.R. are

1) Data preprocessing step
2) fitting logistic regression to the training set.
3) predicting the test results.
   a) Test accuracy of the result.
   b) Visualizing the test set result.

LR can be used where target variable is continous in nature for classification problem sigmoid function can be applied to LR.

So sigmoid function converts the value of y between 0 & 1.



3) What is confusion matrix and how to use it.

→ A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand but the related terminology can be confusing.

Now we have 2X2 matrix -

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

Target Variable - Positive or Negative
Columns - Actual Values
Rows - Predicted Values.

We can calculate accuracy, presicon & recall from Confusion matrix.

4) What is accuracy, Precision, Recall, F-measure.

→ ① Accuracy -

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

② Precision -

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

③ Recall -

Recall is the ratio of correctly predicted positive observations in actual class.

④ F-measure -

F-measure is the weighted average of Precision & recall.