**Aim:**

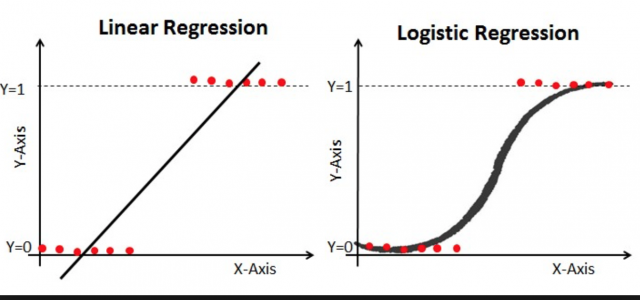
To visualize the data and find the logistic regression application in the dataset

To implement logistic regression using scikit learn

**Theory:**

**What is Logistic Regression?**

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Sometimes logistic regressions are difficult to interpret; the Intellectus statistics tool easily allows you to conduct the analysis, then in plain English interprets the output.



**Why not linear regression?**

1. If you use linear regression, the predicted values will become greater than one and less than zero if you move far enough on the X-axis. Such values are theoretically inadmissible.
2. One of the assumptions of regression is that the variance of Y is constant across values of X (homoscedasticity). This cannot be the case with a binary variable, because the variance is PQ. When 50 percent of the people are 1s, then the variance is .25, its maximum value. As we move to more extreme values, the variance decreases. When P=.10, the variance is .1\*.9 = .09, so as P approaches 1 or zero, the variance approaches zero.
3. The significance testing of the b weights rest upon the assumption that errors of prediction (Y-Y') are normally distributed. Because Y only takes the values 0 and 1, this assumption is pretty hard to justify, even approximately. Therefore, the tests of the regression weights are suspect if you use linear regression with a binary DV.

**Experiment:**

**Logistic Regression Problem Definition:**

1. Predict the accuracy with which you have predicted whether a person has survived or not.

2. Predict whether the flower is Setosa or not using input parameters.

**Colab Link:**

<https://colab.research.google.com/drive/1syJK2zik_n0O5MmGhKEuGzGWigUI2_QE?usp=sharing>

**Conclusion:** The data analysis was performed on IRIS dataset and the following results were identified:

* Using Petal Length and Petal Width we can identify whether the flower is Setosa or not with 100% Accuracy
* Accuracy of prediction of survived cases in Titanic data is 80%