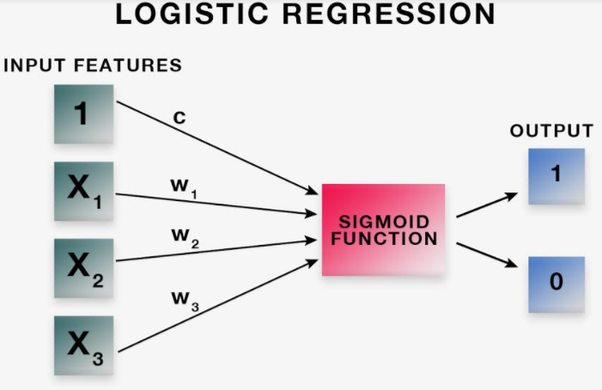
## Logistic Regression

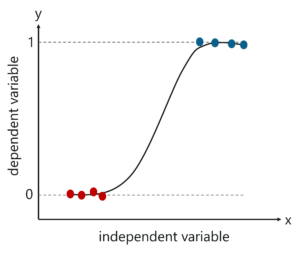
Classification techniques are an essential part of machine learning and data mining applications. Approximately 70% of problems in Data Science are classification problems. There are lots of classification problems that are available, but the logistics regression is common and is a useful regression method for solving the binary classification problem.

Basically, logistic regression model is one of the members of supervised classification algorithm family. Logistic regression measures the relationship between dependent variables and independent variables by estimating the probabilities using a logistic function.

Here, if we talk about dependent and independent variables then dependent variable is the target class variable we are going to predict and on the other side the independent variables are the features we are going to use to predict the target class.

In logistic regression, estimating the probabilities means to predict the likelihood occurrence of the event. For example, the shop owner would like to predict the customer who entered into the shop will buy the play station (for example) or not. There would be many features of customer − gender, age, etc. which would be observed by the shop keeper to predict the likelihood occurrence, i.e., buying a play station or not. The logistic function is the sigmoid curve that is used to build the function with various parameters.

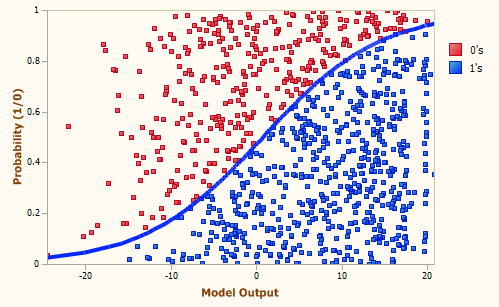




# Logistic Regression Assumptions

* Binary logistic regression requires the dependent variable to be binary.
* For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
* Only the meaningful variables should be included.
* The independent variables should be independent of each other. That is, the model should have little or no multicollinearity.
* The independent variables are linearly related to the log odds.
* Logistic regression requires quite large sample sizes.

Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is dichotomous in nature. Dichotomous means there are only two possible classes. The real life example of classification example would be, to categorize the mail as spam or not spam, to categorize the tumor as malignant or benign and to categorize the transaction as fraudulent or genuine. All these problem’s answers are in categorical form i.e. Yes or No. and that is why they are two class classification problems.



# Training and Predicting

from sklearn.linear\_model import LogisticRegression

logmodel = LogisticRegression()  
logmodel.fit(X\_train,y\_train)

predictions = logmodel.predict(X\_test)

Logistic regression has an array of applications. Here are a few applications used in real-world situations.

**Marketing**: A marketing consultant wants to predict if the subsidiary of his company will make profit, loss or just break even depending on the characteristic of the subsidiary operations.

**Human Resources**: The HR manager of a company wants to predict the absenteeism pattern of his employees based on their individual characteristic.

**Finance**: A bank wants to predict if his customers would default based on the previous transactions and history.