B.Tech CSE (ML)

**Machine Learning Application**

[Fake News Detection](https://www.canva.com/design/DAEy6o35yHk/mxPZcETDJXQkaEscWlY7nw/view?utm_content=DAEy6o35yHk&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink)

horizontal line

# 

# Introduction

Fake news is false or misleading information presented as news. It often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue. However, the term does not have a fixed definition, and has been applied more broadly to include any type of false information, including unintentional and unconscious mechanisms, and also by high-profile individuals to apply to any news unfavorable to their personal perspectives.

## 

## Fake News

False news, also known as unwanted news, false stories, misconceptions or fraudulent stories, types of stories that contain deliberate information or frauds that are spread through traditional media (print and broadcast) or online media. Digital news has revived and increased the use of fake news, or yellow journalism. These stories are often referred to as information that is not in the media but sometimes finds its way into the mainstream media as well.

\* It causes panic

\* Damaging the reputation of public and private organizations

\* It deceives the people, for the benefit of the deceivers

\* Motivated by a personal vendetta, some people support such things.

**PROBLEM :**

How to distinguish between real news and fake news?

**SOLUTION :**

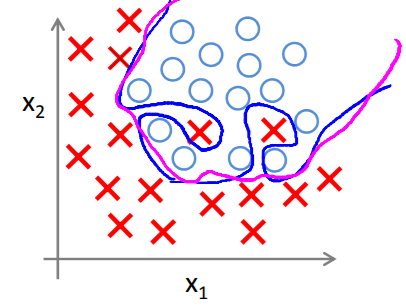
We can show an algorithm that has a large number of false and real news stories to learn to distinguish between them automatically, and then give a possible score or percentage of confidence such as whether a given news release is true or false.

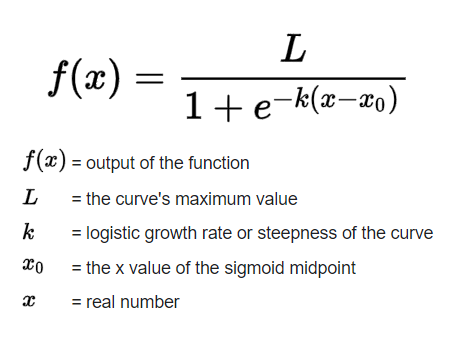
**Methodology followed:**

* Logistic regression classification :
* Support Vector classification :
* Naive-Bayes classification :

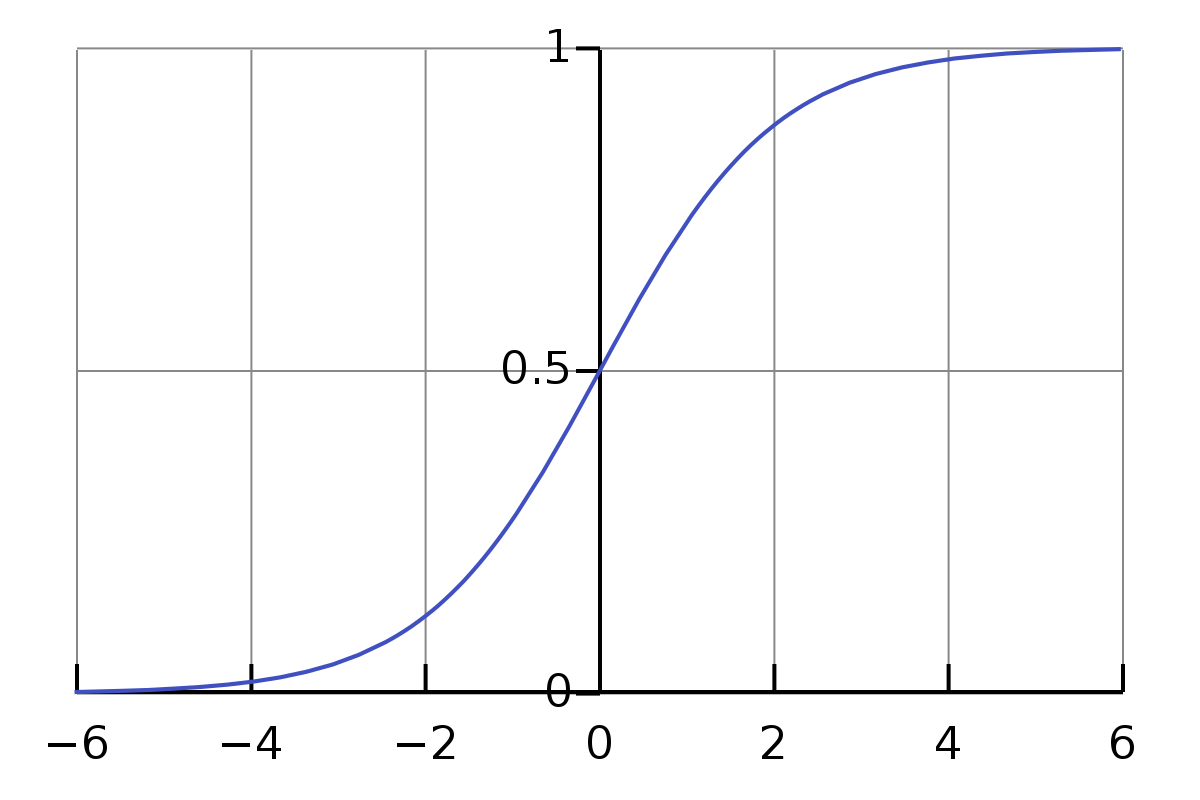
**Logistic Regression**

* It’s a type of classification algorithm that is applied when the label y (output) is a discrete value.
* Classification is a process related to categorization, the process in which ideas and objects are recognized, differentiated and understood

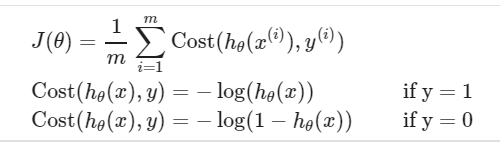


* Logistic Function aka "Sigmoid Function"

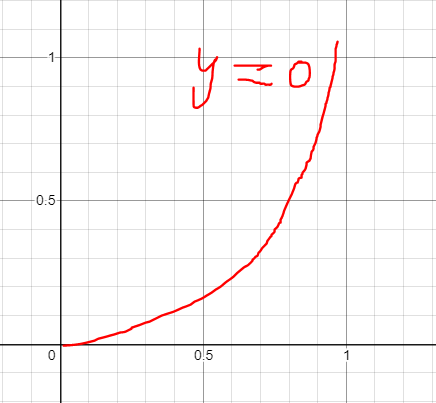
The Graph for Sigmoid function looks like :



* Logistic Function causes the output to be wavy, causing many local optima. In other words, it will not be a convex function, so the cost function for Logistic regression looks like :



If our correct answer 'y' is 0, then the cost function will be 0 if our hypothesis function also outputs 0. If our hypothesis approaches 1, then the cost function will approach infinity.



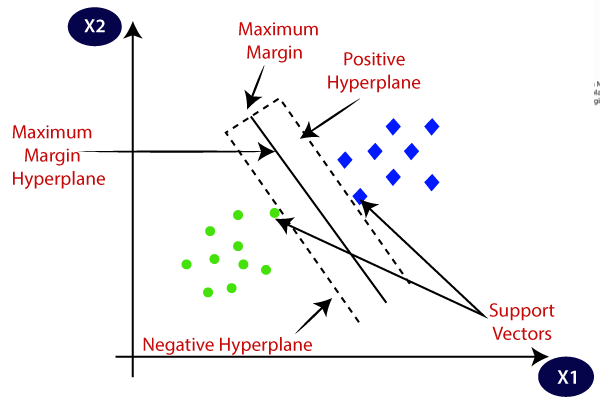
If our correct answer 'y' is 1, then the cost function will be 0 if our hypothesis function outputs 1. If our hypothesis approaches 0, then the cost function will approach infinity.

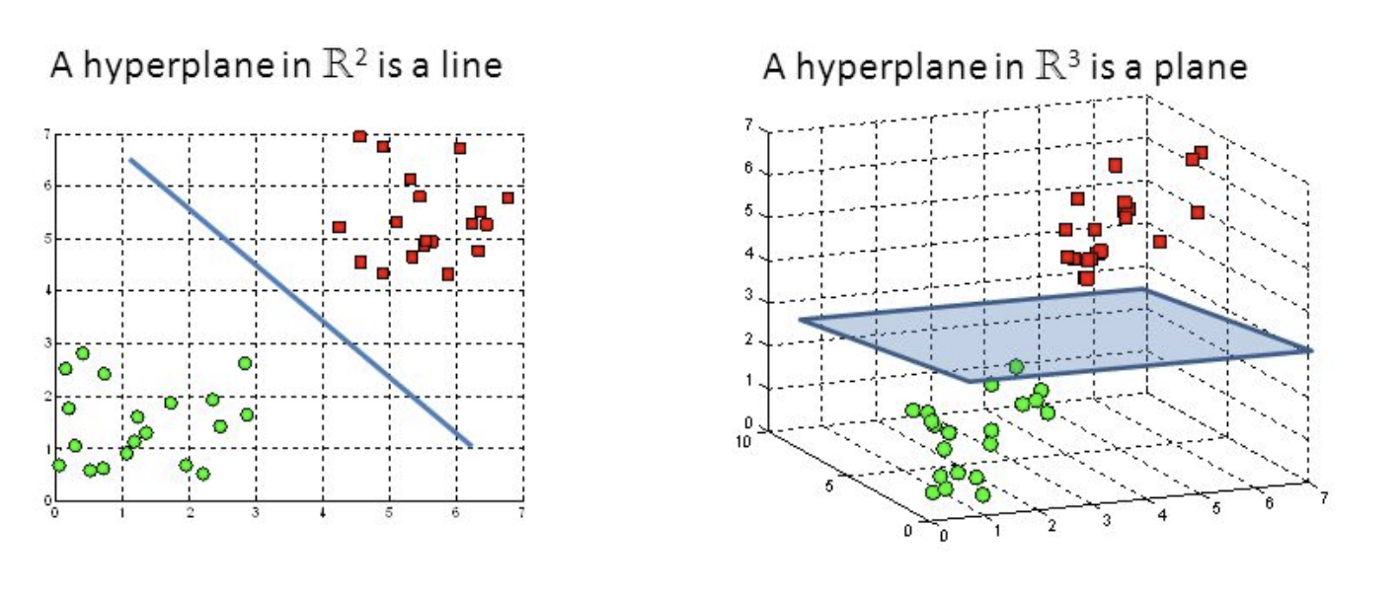


## Support vector machine

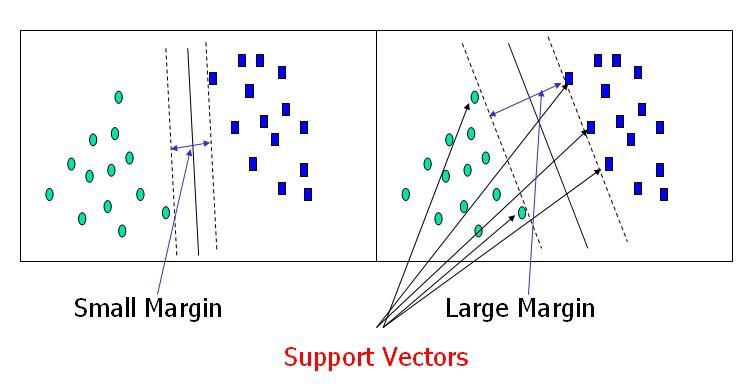
An SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

SVM looks at the extremes of the data and draws a decision boundary also known as a hyperplane near the extreme points of the dataset. So essentially the SVM is a frontier which best segregates the two classes.

****

**Hyperplane** : Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. *If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.*

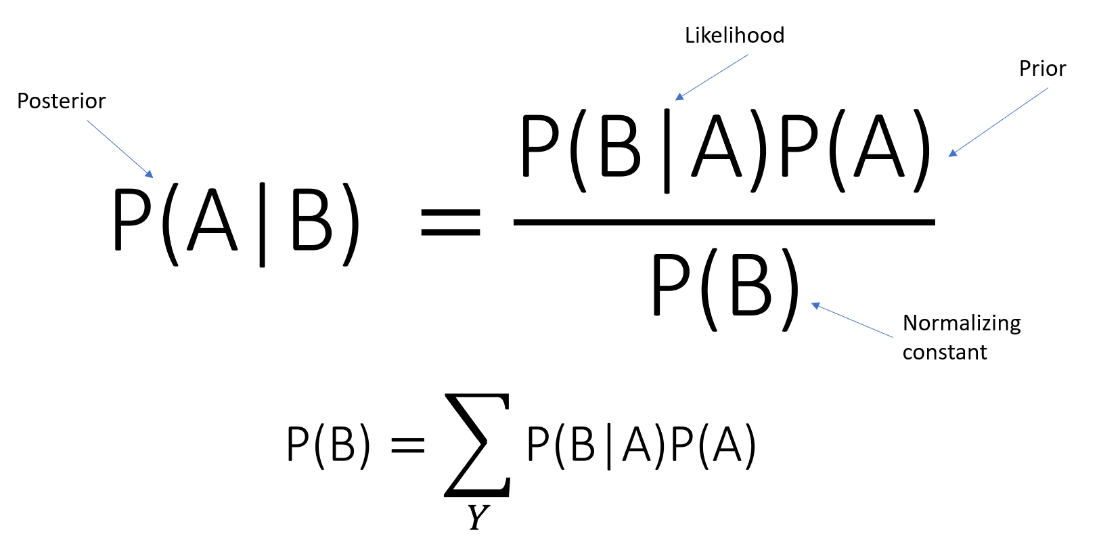
**Support Vector :** Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.



## Naive Bayes Algorithm

Naive Bayes Algorithm is a type of classification of algorithms. It works on Bayes theorem of probability to predict the class of unknown data sets

Bayes theorem can be formulated as :

.

**Prior =>** it describes the degree to which we believe the model accurately describes reality based on all prior information.

**Likelihood =>** it describes how well the model predicts the data.

**Normalizing constant =>** The constant that makes the posterior density integrate to one.

**Posterior =>** represents the degree to which we believe a given model accurately describes a given model accurately describes the situation given the available data and all or our prior information.