

# INTRODUCTION

# CHAPTER 1

## INTRODUCTION

### 1.1 INTRODUCTION TO ASD

The problem of autism spectrum disorder (ASD) have been mounting swiftly nowadays among all ages of the human population. Early detection of this neurological disease can greatly assist in the maintenance of the subject's mental and physical health. With the rise of application of machine learning-based models in the predictions of various human diseases, their early detection based on various health and physiological parameter now seems possible. This factor motivated us to increase interest in the detection and analysis of ASD diseases to improve better treatment methodology. Detection of ASD becomes a challenge as there are several other mental disorders whose few symptoms are very similar to those with ASD symptoms, thereby makes this task a difficult one.

Autism Spectrum disorder is a problem that is related to human brain development. A person who has suffered from the Autism Spectrum Disorder is generally not able to do social interaction and communication with other persons. In this, a person's life is usually affected for his or her entire lifetime. It is interesting to know that both environmental and genetic factors may turn out to be the causing factors for this disease. The symptoms of this problem may be started at the age of three years and may continue for the lifetime. It is not possible to complete treat the patient suffering from this disease, however its effects can be reduced for some time if the symptoms are early detected. By assuming that human genes are responsible for it, the exact causes of ASD have not been recognized by the scientist yet. The human genes affect the development by influencing the environment. There is some risk factor which influences ASD like as low birth weight, a sibling with ASD and having old parents, etc.

## 1.2 INTRODUCTION TO MACHINE LEARNING

Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start discussing about the algorithms which we are using.

### 1.2.1 Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are

called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

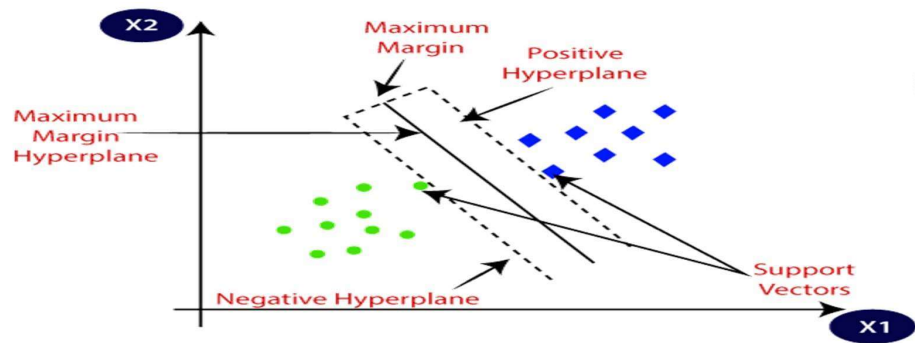


Fig 1.1 SVM

### 1.2.2 NAIVE BAYES

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

$$P(A|B) = (P(B|A) \cdot P(A)) / P(B)$$

### 1.2.3 Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of

independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

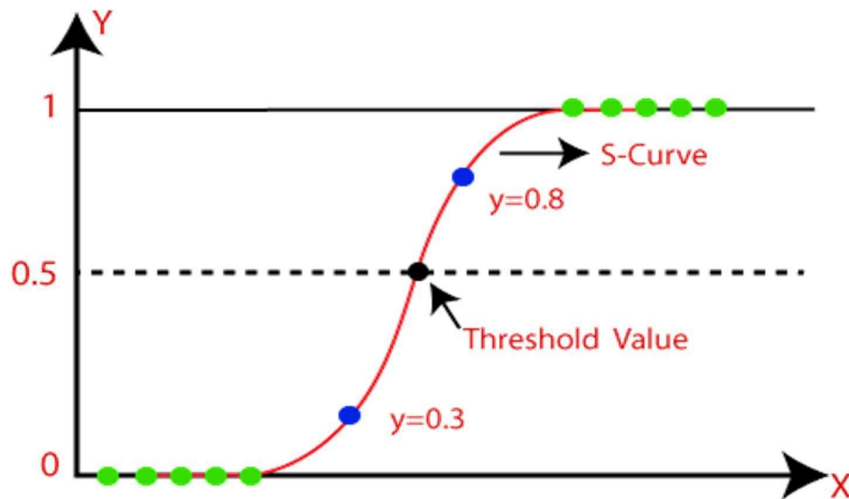


Fig 1.2 Logistic regression

#### 1.2.4 K-Nearest Neighbors Algorithm

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

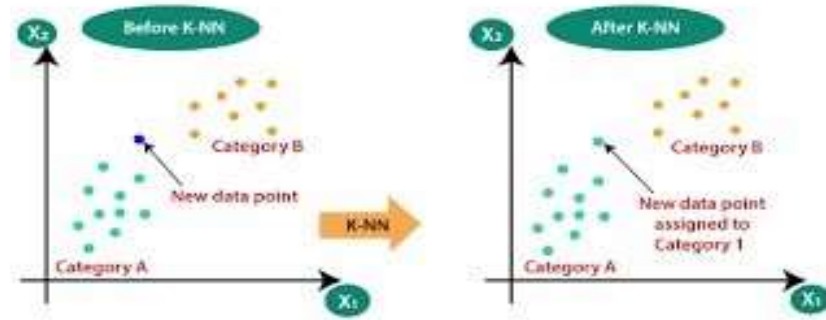


Fig 1.3 KNN

### 1.2.5 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information.

### 1.2.6 Multilayer perceptron

Multilayer perceptron (MLP) is a supplement of feed forward neural network. It consists of three types of layers—the input layer, output layer and hidden layer. The input layer receives the input signal to be processed. The required task such as prediction and classification are performed by the output layer. An arbitrary number of hidden layers that are placed in between the input and output layer are the true computational engine of the MLP.

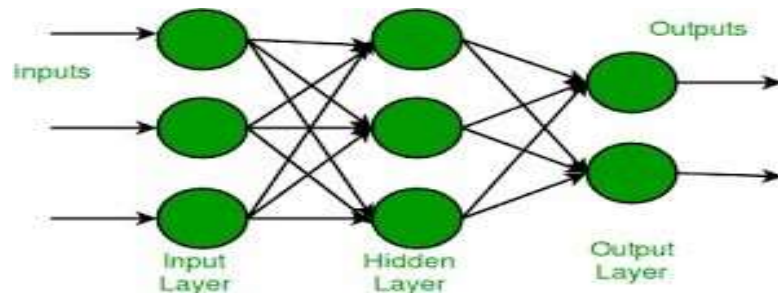


FIG 1.4 MLP

### 1.3 OBJECTIVE

Autistic Spectrum Disorder (ASD) is the name for a group of developmental disorders impacting the nervous system. ASD symptoms range from mild to severe: mainly language impairment, challenges in social interaction, and repetitive behaviors. Many other possible symptoms include anxiety, mood disorders and Attention-Deficit/Hyperactivity Disorder (ADHD). ASD has a significant economic impact in the healthcare domain, both due to the increase in the number of ASD cases, and because of the time and costs involved in diagnosing a patient. Early detection of ASD can help both patients and the healthcare sector by prescribing patients the therapy and/or medication they need and thereby reducing the long-term costs associated with delayed diagnosis. Thus, health care professionals across the globe have an urgent need for easy, time-efficient, robust and accessible ASD screening methods that can accurately predict whether a patient with certain measured characteristics has ASD and inform individuals whether they should pursue formal clinical diagnosis. However, challenges remain. Pursuing such research necessitates working with datasets that record information related to behavioural traits and other factors such as gender, age, ethnicity, etc. Such datasets are rare, making it difficult to perform thorough analyses to improve the efficiency, sensitivity, specificity and predictive accuracy of the ASD screening process. At present, very limited autism datasets associated with clinical or screening are available and most of them are genetic in nature. These data are extremely sensitive and hard to collect for social and personal reasons and the regulations around them.

## 1.4 PROBLEM STATEMENT

In the world there are so many neural disorders but ASD can be cured with prior treatment. A new study published in Autism Research shows that about 100 in 10,000 (or 1 in 100) children around the world are diagnosed with autism spectrum disorder (ASD). This is an increase from the 2012 global prevalence report which found that 62 in 10,000 children were autistic. Autism Spectrum Disorder will affect the social life like impaired communication, impaired reciprocal social interaction and restricted, repetitive and stereotyped patterns of behaviors or interests.