Prior Knowledge

Date	26 September 2022
Team ID	PNT2022TMID52603
Project Name	A Novel Method for Handwritten Digit Recognition System

One should have knowledge on the following Concepts:

1. MACHINE LEARNING:

Machine learning is the science of making computers learn and act like humans by feeding data and information without being explicitly programmed.

Types of Machine Learning:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Reinforcement Learning

Semi-Supervised learning is used in Text Classification.

1. **Supervised Learning:**

In supervised learning, a model is able to predict with the help of labeled dataset. It is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.

Types:

a) Classification

b) Regression

a) Classification:

Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "**Yes" or No, Male or Female, Red or Blue, etc**. The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are **Spam Detection, Email filtering, etc.**

b) Regression:

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

2. Unsupervised learning:

Unsupervised Learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision. It means, in unsupervised machine learning, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.

Types:

a) clustering

b) Association

a) Clustering:

The method of dividing the objects into clusters which are similar between them and are dissimilar to the objects belonging to another cluster.

b) Association:

Discovering the probability of the co-occurrence of items in a collection.

2.CLUSTERING, CLASSIFICATION AND REGRESSION

Clustering:

Clustering is an unsupervised technique. With clustering, the algorithm tries to find a pattern in data sets without labels associated with it. This could be a clustering of buying behavior of customers. Features for this would be the household income, age and clusters of different consumers could then be built.

Classification:

In contrast to clustering, classification is a supervised technique. Classification algorithms look at existing data and predicts what a new data belongs to. Classification is used for spam for years now and these algorithms are more or less mature in classifying something as spam or not. With machine data, it could be used to predict a material quality by several known parameters such as humidity, strength, color, etc. The output of the material prediction would then be the quality.

Regression:

Regression is often confused with clustering, but it is still different from it. With a regression, no classified labels (such as good or bad, spam or not spam, ...) are predicted. Instead, regression outputs continuous, often unbound, numbers. This makes it useful for financial prediction.

3.Artificial Neural Networks:

Artificial neural networks, usually simply called neural networks or neural nets, are computing systems inspired by the biological neural networks that constitute animal brains. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

4.Convolutional Neural Networks:

A convolutional neural network (CNN) is a type of artificial neural network used primarily for image recognition and processing, due to its ability to recognize patterns in images. A CNN is a powerful tool but requires millions of labelled data points for training.

Components of CNN:

1) Input layer
2) Output layer
3) One or more hidden layer