Chapter 1

General principles of machine learning

1-1-Artificial intelligence and machine learning

From about sixty years ago, the growth of data to evaluate a phenomenon was exponentially formed. The amount of this growth is very high and now it is beyond human intelligence and in order to process it, in addition to traditional computer programs, it is necessary to use equipment and artificial intelligence based on neural networks.

This approach simulates the learning processes of the human brain; hence it is called Machine Learning (ML).

The traditional method of data processing is based on an algorithm and the computer has a static mode and software to run a specific program, which is shown in Figure 1 (a). In this approach, the program is customized with specific data to be processed. Optimizing this process relies on humans to update the program.

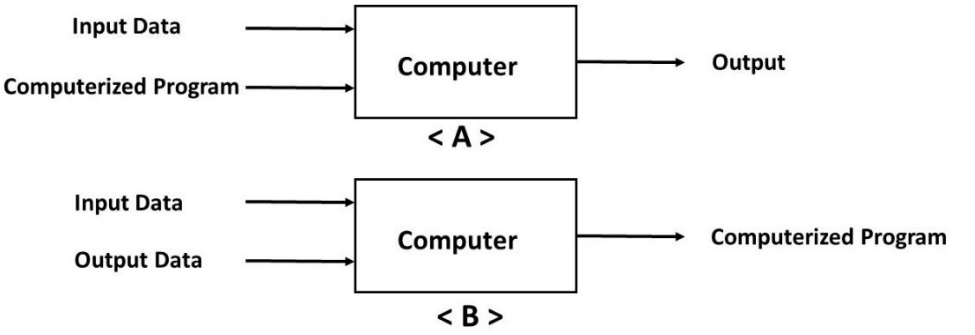

Figure 1) Traditional programming schemes (a) versus machine learning (b).

Machine learning, on the other hand, has a different design, as shown in Figure 1 (b). First, one must design the neural network configuration and adjust the number of its parameters.

Second, in a supervised training process, the system is fed with both the correct input and output data, and the system is dynamically optimized by comparing them, that is, the output is calculated with the correct items presented at the beginning. With adequate training processes, the system can process new data with high reliability.

The first part of the neural network production is called the Perceptron, which is shown in Figure 2. Perceptron has three layers of artificial neurons: an input layer, a hidden layer and an output layer. In a dense perceptron, each neuron in the latent layer is connected to all the nerve cells in the input and output layers.

The input layer data are aggregated together with different weights and are associated with hidden layer neurons. There is also a connection between the hidden layer and the output layer. These weights are the key parameters for optimization in the training process. Initially, these weights can be determined by random numbers, but according to the researcher's initial information, there are advanced approaches to initial quantification.

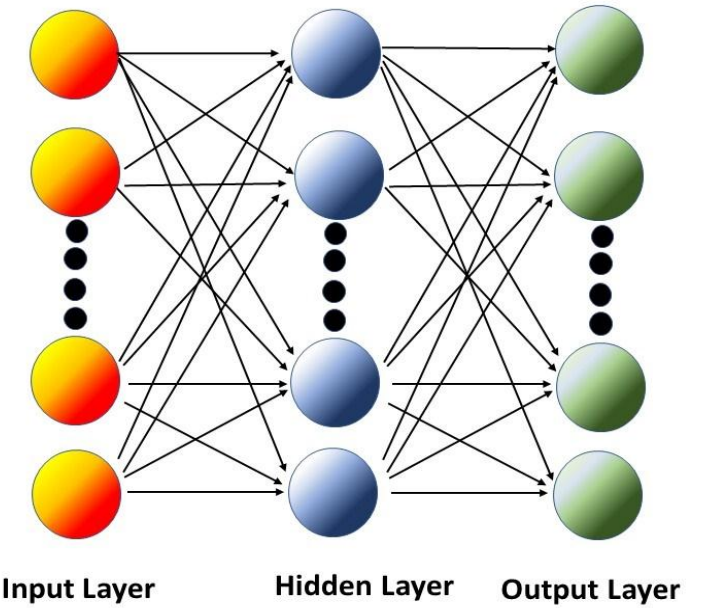

Figure 2) Arrange the layers of a neural network

With a hidden layer, the ability of perceptron is relatively limited. Natural evolution is the addition of more hidden layers, which is also reflected in biological evolution. Turtles, for example, have three layers of neurons, but the human brain has six layers in the neocortex. Architecture with many hidden layers is called Deep Neural Network (DNN) and its function is very precise and deep, hence it is called Deep Learning (DL).

One of the main uses of ML is in image recognition, where the data is in 2D format.

If a traditional neural network is used, the matrix will be two-dimensional. In one-dimensional conditions the output is vector

One-dimensional and flat. In two-dimensional form, there is a strong correlation between neighboring pixels, but this Important information is complex in one dimension.

Thus, a new approach was developed called the Neural Network Convolutional (CNN) (Figure 3). On a typical CNN, hidden layers also have a two-dimensional matrix called filters.

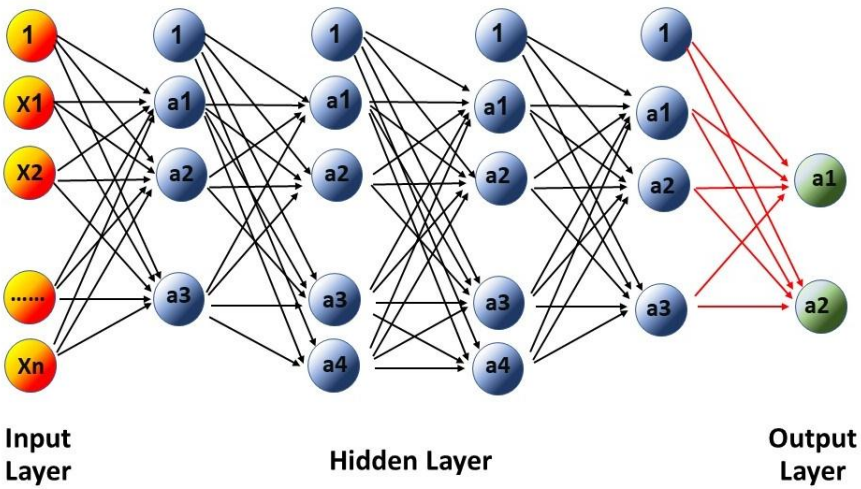

Figure 3) Pattern of a complex neural network or convolution

The CNN architecture is something like the human cerebral cortex, in that the occipital lobe of the human brain receives and processes retinal information in several stages in the primary visual cortex. Initially, only the initial patterns, such as the vertical and horizontal edges of an object, are identified. These are the results of the primary visual cortex. In the secondary area, etc., significant progress is being made in image recognition with

CNNs. At present, the detection of cancer cells through CT images has gone beyond human function in many areas.

In addition to image recognition, neural networks can also translate languages or analyze sound. This information, which is the function of the temporal lobe of the human brain, carries sound signals with time series information, unlike the information in an image, and so memory is closely involved here. An audio signal can be recorded in a one-dimensional vector, and there is no mention of two-dimensional correlation in it, but on the other hand, there is a temporal correlation in the signal.

For example, one word in a sentence often refers to another word in the previous sentence. Therefore, neural networks require a special configuration to process audio information, and the common type is called the Recurrent Neural Network (RNN), as shown in Figure 4.

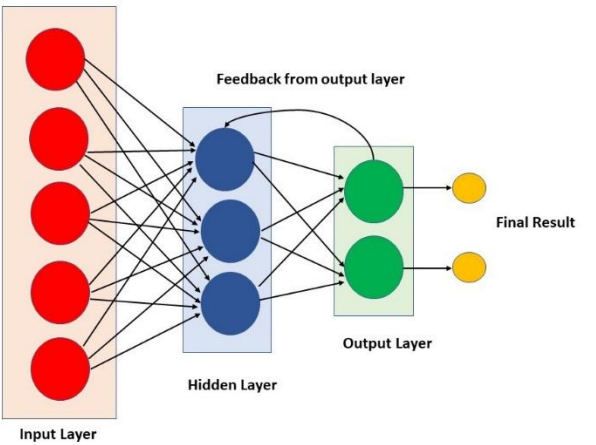

Figure 4) Hidden layer feedback from the output layer for correction and adaptation and network self-regulation.