Chapter 5

INTRODUCTION TO ARTIFICIAL NEURAL NETWORK

Artificial neural networks or simply “neural nets” go by many names such as connectionist models, parallel distributed processing models, and neuromorphic systems. Whatever terminology it may be, they all attempt to borrow the structure and running way of the biological nervous system based on our present understanding of it. Instead of performing a program consisting of instructions sequentially as in a von Neumann computer, artificial neural nets have their structures in dense interconnection of simple computational elements—the artificial neurons or simply “neurons”, and operate the massive computational elements in parallel to achieve high performance speed.

Neurons in a neural net can be viewed as nodes in a layer network defined in Section 1.5, but as a node in a neural net the neuron not only sums up the weighted inputs from other nodes in one of the neighboring layers but also performs a nonlinear transformation on the summation, then the output of this neuron will be sent to all neurons in the next layer with links to it (in a network in graph theory a node distributes its received input to nodes linked). Nodes or computational elements in neural nets are nonlinear and typically realized by analog circuits. Different types of nodes, distinguished by types of nonlinearities, can be used in one network. So there are three key factors for specifying a neural net:

- The net topology. Topological factors include: feedforward type network or feedback type network, the number of layers, and the number of nodes in each layer.

- The type of nodes (neurons). Different nonlinearities realized by analog circuits or more complex mathematical operations realized by digital circuitries can be considered. The type of neurons also determines the time
feature of the network operation: the nodes operate continuously or at
discrete amounts of time.

- The weights specification. There are two cases: predetermined weights and
adapted weights. Adaptation or learning is the main feature of artificial
neural nets. The ability to adapt and continue learning is essential in areas
such as speech recognition.

Classification of neural nets then can be made on the basis of the above three
factors. Considering the topological structure, there are two classes of neural
nets: feedforward neural nets and feedback neural nets. Nets introduced
by Hopfield ([148], [150], [152]) are in the class of feedback nets. A class
of artificial neural networks called perceptrons ([261], [225]) introduced by
Rosenblatt has the feedforward structure.

Different nonlinearities of the neurons specify the features of input and
output of the nets. A binary net is a net with binary input and output. These
nets are most appropriate in processing images which have pixel values 1 or −1
(black and white image). A continuous net is a net with continuous input and
output. There could be hybrid nets which have binary input and continuous
output, or continuous input and binary output.

Training strategy is also a key to classify the artificial neural nets. An
adaptive net is a net whose weights are trained during the operation by learning
or self-adaptation. Nets trained by self-organization or self-adaptation (self-
organized nets) are also said to be nets trained without supervision. As we
will see, Kohonen’s feature-map forming nets ([180], [183]) are typical nets
without supervision. Nets trained by learning are said to be nets trained with
supervision. Nets with fixed weights are non-adaptive nets, and most of the
artificial neural nets that were designed for solving mathematical programming
problems are non-adaptive nets as we will see in the following chapters.

5.1 WHAT IS AN ARTIFICIAL NEURON?

Mathematically, an artificial neuron is a composite nonlinear function \( a(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) (shortly we call \( a(x) \) the neuron function). More precisely,

\[
a(x) = \phi(\sigma(x) - T),
\]

where \( T \in \mathbb{R} \) is the external threshold or simply threshold, \( \sigma(x) : \mathbb{R}^n \rightarrow \mathbb{R} \)
is called an accumulation function. And \( \phi(x) : \mathbb{R} \rightarrow \mathbb{R} \) is a nonlinear
function called activation function. The accumulation function is generally a
linear function of the input \( x \). In this case,

\[
a(x) = \phi(w^T x - T),
\]