Chapter 5

ARTIFICIAL NEURAL NETWORK

Inspired by the sophisticated functionality of human brains where hundreds of billions of interconnected neurons process information in parallel, researchers have successfully tried demonstrating certain levels of intelligence on silicon. Examples include language translation and pattern recognition software. While simulation of human consciousness and emotion is still in the realm of science fiction, we, in this chapter, consider artificial neural networks as universal function approximators. Especially, we introduce neural networks which are suited for time series forecasts.

5.1 Introduction

An artificial neural network (or simply neural network) consists of an input layer of neurons (or nodes, units), one or two (or even three) hidden layers of neurons, and a final layer of output neurons. Figure 5.1 shows a typical architecture, where lines connecting neurons are also shown. Each connection is associated with a numeric number called weight. The output, $h_i$, of neuron $i$ in the hidden layer is,

$$ h_i = \sigma\left(\sum_{j=1}^{N} V_{ij} x_j + T_i^{\text{hid}}\right), \quad (5.1) $$

where $\sigma()$ is called activation (or transfer) function, $N$ the number of input neurons, $V_{ij}$ the weights, $x_j$ inputs to the input neurons, and $T_i^{\text{hid}}$ the threshold terms of the hidden neurons. The purpose of the activation function is, besides introducing nonlinearity into the neural network, to bound the value of the neuron so that the neural network is not paralyzed by divergent neurons. A common example of the activation function is the sigmoid (or logistic) function.
defined as (Figure 5.2),

$$\sigma(u) = \frac{1}{1 + \exp(-u)}.$$  \hspace{1cm} (5.2)

Other possible activation functions are arc tangent and hyperbolic tangent. They have similar response to the inputs as the sigmoid function, but differ in the output ranges.

It has been shown that a neural network constructed the way above can approximate any computable function to an arbitrary precision. Numbers given to the input neurons are independent variables and those returned from the output neurons are dependent variables to the function being approximated by the neural network. Inputs to and outputs from a neural network can be binary (such as yes or no) or even symbols (green, red, ...) when data are appropriately encoded. This feature confers a wide range of applicability to neural networks.