Chapter 1
Introduction

Building a machine or autonomous mechanism endowed with intelligence is an ancient dream of researchers from the diverse areas of sciences and engineering. Although the first articles about Artificial Neural Networks (ANN) were published more than 50 years ago, this subject began to be deeply researched on the early 90s, and still have an enormous research potential. The applications involving systems considered intelligent cover a wide range, including:

- Analysis of images acquired from artificial satellites.
- Speech and writing pattern classification.
- Face recognition with computer vision.
- Control of high-speed trains.
- Stocks forecasting on financial market.
- Anomaly identification on medical images.
- Automatic identification of credit profiles for clients of financial institutions.
- Control of electronic devices and appliances, such as washing machines, microwave ovens, freezers, coffee machines, frying machines, video cameras, and so on.

Besides these applications, the capabilities of artificial neural networks enable the solution of other kinds of problems deriving from different knowledge areas, as testified on numerous scientific journals. One example is medicine, where artificial neural networks are used for classifying and predicting cancer based on the genetic profile of a given individual (Cireşan et al. 2013; Khan et al. 2001). Other application, presented by Yan et al. (2006), proposes a decision support system, also based on artificial neural networks, for heart disease diagnosis.

The chemistry area has also registered works where artificial neural networks are used for obtaining novel polymeric compounds (Steinera et al. 2011; Zhang and Friedrich 2003). Artificial neural networks are also applied in control systems for water treatment, which involves physical and chemical nonlinear processes that are difficult to be mapped by conventional control methods (Han et al. 2011; Zhang and Stanley 1999).

© Springer International Publishing Switzerland 2017

I.N. da Silva et al., *Artificial Neural Networks*,
DOI 10.1007/978-3-319-43162-8_1

On the biology area, it is possible to find applications using artificial neural networks with the goal of identifying bat species based on their echo localization signals (biosonar) emitted during flight (Armitage and Ober 2010; Parsons and Jones 2000). Another neural approach for classifying mice species, through the sound they produced, was developed by Tian and Shang (2006).

Concerning financial and economic fields, there are also problems that are difficult to solve, mostly due to the non-linear behavior of these systems. Artificial neural networks are widely applied to such scenarios thanks to their capabilities of handling intrinsically nonlinearities (Karaa et al. 2011; Coakley and Brown 2000).

The biology field has also been beneficiated from the capabilities of artificial neural networks for extracting information, and uses them to analyze the influence of the weather on the growing of trees (Lek and Guégan 2012; Zhang et al. 2000).

The abilities of artificial neural networks for pattern classification can be observed even in the etiology field, where they are used to distinguish the diverse facial expressions conveying human emotions (Agarwal et al. 2010; Dailey and Cottrell 2002). Another application related to pattern classification is discussed by Fernandes et al. (2013), where artificial neural networks are used to classify harmonic current sources in power distribution systems.

It is possible to find in the pharmaceutical field the employment of artificial neural networks to support the formulation of novel drugs, indicating whether the medicine should be produced by microemulsion or solid dispersion methods (Deeb 2010; Mendyk and Jachowicz 2007).

On acoustics, it is also possible to find research that uses artificial neural networks for assaying the environmental acoustic impedance, a very important feature for projecting cinema rooms and environments sensitive to external noise (Hinton et al. 2012; Too et al. 2007).

The depth in which pollutants are expected to penetrate in the soil and contaminate ground water can also be estimated through artificial neural networks, providing some basis for the development of contention actions (Chowdhury et al. 2010; Tabach et al. 2007).

The food industry has also been beneficiated from the application of artificial neural networks (Argyria et al. 2010), such as those applied in classifying different varieties of tea (He et al. 2007). Another application example is presented by Silva (2007), where a neural approach was developed for processing signals from nuclear magnet resonance to classify cattle beef, allowing the identification of the sex and race of the animals. On a different application, Nazarion (2007) employed artificial neural networks combined with ultrasound techniques to classify milk regarding adulteration by the addition of fat and water.

In the automotive and aerospace industry, applications with artificial neural networks are found for assisting the mapping of processes involving estimation of control variables and project parameters. As an example of such applications, Cho et al. (2006) proposed modeling methods and control strategies for unmanned aerial vehicles. Richter et al. (2010) designed a neural architecture to perform virtual sensing of oxygen in bi-fuel vehicles. Vicente et al. (2007) proposed a neural idling speed controller for internal combustion engines. Another interesting application in the automotive area is established by Ortega and Silva (2008), where artificial neural networks are used to optimize brake light projects built with light emitting diodes.

Artificial neural networks are part of the area known as intelligent systems (connectionist systems), or computational intelligence (Jang et al. 1997; Zadeh 1992). Besides artificial neural networks, the intelligent system area includes diverse tools, such as fuzzy systems, (Pedrycz and Gomide 2007; Buckley and Siler 2004; Ross 2004), evolutionary computing (Michalewicz 1999; Dasgupta and Michalewicz 1997; Goldberg 1989), swarm intelligence (Kennedy and Eberhart 2001), artificial immunologic systems (Dasgupta 2006; Castro and Timmis 2002) and intelligent agents (D'Inverno and Luck 2004).

Additionally, today's entertainment industry, especially the cinematographic arts, has explored the subject in many science fiction movies that address the use of intelligent systems in machines and robots.

The most attractive feature of artificial neural networks, and also the source of their reputation as powerful tools for solving diverse problems, is their high capacity of mapping nonlinear systems, enabling them to learn the underlying behaviors from data acquired from such systems.

1.1 Fundamental Theory

Artificial neural networks are computational models inspired by the nervous system of living beings. They have the ability to acquire and maintain knowledge (information based) and can be defined as a set of processing units, represented by artificial neurons, interlinked by a lot of interconnections (*artificial synapses*), implemented by vectors and matrices of synaptic weights.

1.1.1 Key Features

The most relevant features concerning artificial neural applications are the following:

(a) *Adapting from experience*
The internal parameters of the network, usually its synaptic weights, are adjusted with the examination of successive examples (patterns, samples, or measurements) related to the process behavior, thus enabling the acquisition of knowledge by experience.

(b) *Learning capability*
Through the usage of a learning method, the network can extract the existing relationship between the several variables of the application.

(c) *Generalization capability*
Once the learning process is completed, the network can generalize the acquired knowledge, enabling the estimation of solutions so far unknown.
(d) *Data organization*
Based on innate information of a particular process, the network can organize this information, therefore enabling the clustering of patterns with common characteristics.
(e) *Fault tolerance*
Thanks to the high number of interconnections between artificial neurons, the neural network becomes a fault-tolerant system if part of its internal structure is corrupted to some degree.
(f) *Distributed storage*
The knowledge about the behavior of a particular process learned by a neural network is stored in each one of the several synapses between the artificial neurons, therefore improving the architecture robustness in case of some neurons are lost.
(g) *Facilitated prototyping*
Depending on the application particularities, most neural architectures can be easily prototyped on hardware or software, since its results, after the training process, are usually obtained with some fundamental mathematical operations.

1.1.2 Historical Overview

The first publication related to neurocomputing dates from 1943, when McCulloch and Pitts (1943) composed the first mathematical model inspired by biological neurons, resulting in the first conception of the artificial neuron.

In 1949, the first method for training artificial neural networks was proposed; it was named Hebb's rule and was based on hypothesis and observations of neurophysiologic nature (Hebb 1949).

Many other researchers have continued the development of mathematical models based on the biological neuron, consequently generating a large number of topologies (structures) and learning algorithms. Among the different branches that emerged, the work of Frank Rosenblatt stands out. Between 1957 and 1958, Rosenblatt developed the first neurocomputer called Mark I Perceptron, crafting the basic model of the Perceptron (Rosenblatt 1958).

The Perceptron model stirred interest due to its capability of recognizing simple patterns. Widrow and Hoff (1960) developed a network called ADALINE, which is short for ADAptive LINEar Element. Later on, the MADALINE, the Multiple ADALINE, was proposed. It consisted on a network whose learning is based on the Delta rule, also known as LMS (Least Mean Square) learning method.

Following this earlier work, many researchers of that time were encouraged to conduct research in this area. However, in 1969, neurocomputing suffered a major setback with the publication of the classical book "Perceptrons: An Introduction to Computation Geometry" by Minsky and Papert (1969). The authors discussed emphatically the limitations of the neural networks of that time—which were composed of a single layer, such as the Perceptron and the ADALINE—on learning the relationship between inputs and outputs of very basic logical functions, such as XOR (exclusive or). To be more precise, that book demonstrated the impossibility of neural networks to classify patterns of nonlinearly separable classes.

Following the impact of that publication, researches on neural networks were greatly reduced, and some of the few works thereafter were: the derivation of prediction algorithms using reverse gradients (Werbos 1974), the development of the ART (Adaptive Resonance Theory) network by Grossberg (1980), the formulation of the self-organized maps (SOM) by Kohonen (1982), and the recurrent network based on energy functions proposed by Hopfield (1982). The latter is the work that brought to the artificial neural networks area its original prestige from before 1969.

Only after the end of the 1980s, supported by the work above, scientists restored their interest in this area. The definitive comeback of artificial neural networks is due to different reasons, such as the development of computers with enhanced processing and memory capabilities, the conception of more robust and efficient optimization algorithms, and finally, the novel findings about the biological nervous system. One of the fundamental works of that time was the publication of Rumelhart, Hinton and Williams' book "Parallel Distributed Processing" (Rumelhart et al. 1986), which brought to light one algorithm that allowed the adjustment of weight matrices of networks with more than a single layer. Consequently, solving the old problem of learning patterns from the XOR logical function. The proposal of this algorithm, called "backpropagation," definitely revived and motivated researches in artificial neural networks.

In recent years, together with numerous practical applications on different areas of knowledge, dozens of new and different researches have enabled theoretical advancements in artificial neural networks. Some interesting work, in particular, includes the learning algorithm based on Levenberg–Marquardt method, which fostered efficiency improvement of artificial neural networks in diverse applications (Hagan and Menhaj 1994); the artificial neural networks based on support vector machines (SVM), which can also be used for pattern classification and linear regression (Vapnik 1998); and the development of neural integrated circuits with several circuit configurations (Beiu et al. 2003).

A highly detailed description about the several other historical facts within the evolving process of artificial neural networks, since its early beginning, can be found on Haykin (2009).

1.1.3 Potential Application Areas

Artificial neural networks can be employed in several problems related to engineering and sciences. The potential application areas can be divided as follows:

(a) *Universal curve fitting (function approximation)*
The goal is to map the functional relationship between variables (usually real numbers) of a particular system from a known set of meaningful values. These applications are as diverse as possible, and often involve mapping processes that are difficult to model using traditional methods.

(b) *Process control*
This application category consists of identifying control actions capable of meeting quality, efficiency, and security requirements. Among the multiple available applications, neural controllers are of particular interest to robotics, airplanes, elevators, appliances, satellites, and so on.

(c) *Pattern recognition/classification*
The purpose is to associate a given input pattern (sample) to one of the previously defined classes, as in the case of image, speech and writing recognition. In this case, the problem being addressed has a discrete and known set of possible desired outputs.

(d) *Data clustering*
On this circumstance, the goal is to detect and identify similarities and particularities of the several input patterns to allow their grouping (clustering). Some examples, to cite a few, are applications involving automatic class identification and data mining.

(e) *Prediction system*
The purpose of this system category is to estimate future values of a particular process, taking into account several previous samples observed in its domain. Among the known applications, it is possible to find systems for time series prediction, stock market projection, weather forecast, and so on.

(f) *System optimization*
The goal is to minimize or maximize a cost function (objective) obeying eventual constraints to correctly map a problem. Among the optimization tasks which can benefit from artificial neural networks, the most important includes constrained optimization problems, dynamic programming, and combinational optimization.

(g) *Associative memory*
The objective is to recover a correct pattern even when its inner elements are uncertain or inaccurate. Some examples include image processing, signal transmission, written character identification, and so forth.

1.2 Biological Neuron

The information processing performed by the human brain is carried out by biological processing components, operating in parallel, for producing proper functions, such as thinking and learning.

The fundamental cell of the central nervous system is the neuron, and its role comes down to conduct impulses (electrical stimuli originated from physical–chemical reactions) under certain operation conditions. This biological component can be divided into three main parts: dendrites, cell body (also known as “soma”), and axon.

Dendrites are composed of several thin extensions that form the dendritic tree (Fig. 1.1). The fundamental purpose of dendrites is to acquire, continuously, stimuli from several other neurons (connectors) or from the external environment, which is the case of some neurons in contact with the environment (also called sensory neurons).

The cell body is responsible for processing all the information that comes from the dendrites, to produce an activation potential that indicates if the neuron can trigger an electric impulse along its axon. It is also in the cell body where the main cytoplasmic organelles (nucleus, mitochondria, centriole, lysosome, and so forth) of the neuron can be found.

The axon is composed of a single extension whose mission is to guide the electrical impulses to other connecting neurons, or to neurons directly connected to the muscular tissue (efferent neurons). The axon termination is also composed of branches called synaptic terminals.

The synapses are the connections which enable the transfer of electric axon impulses from a particular neuron to dendrites of other neurons, as illustrated in Fig. 1.2. It is important to note that there is no physical contact between the neurons forming the synaptic junction, so the neurotransmitter elements released on the junction are in charge of weighting the transmission from one neuron to another. In

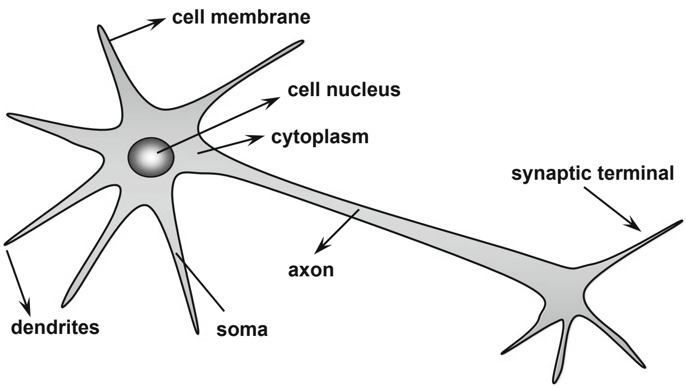


Fig. 1.1 Biological neuron

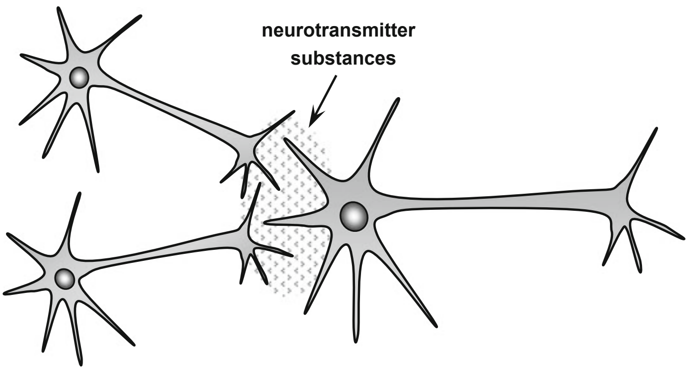


Fig. 1.2 Illustration of the synaptic connection between neurons

Table 1.1 Physical properties of the human brain and its components (in estimated values)

Property	Physical dimension
Brain mass	1.5 kg
Energy consumed by the brain	20 %
Neuron length	100 μm
Resting potential	−70 mV
Threshold potential	−55 mV
Action potential (peak)	35 mV

fact, the functionality of a neuron is dependable of its synaptic weighting, which is also dynamic and dependent on the cerebral chemistry (Hodkin and Huxley 1952).

In short, although the activities related to the biological neuron might seem very simple at first, its components, when functioning altogether, are responsible for all the processing executed and managed by the human brain. It is estimated that this biological neural network, with very eccentric features, is composed of about 100 billion (10^{11}) neurons. Each one of those is interconnected through synaptic connections (made possible by more than fifty neurotransmitter substances) to an average of 6,000 neurons, thus resulting in a total of 600 trillion synapses (Shepherd 2004). Table 1.1 presents some physical properties about the human brain (to be more precise, of an adult human) and its components.

As presented in Table 1.1, the neural membrane action potential has negative values when resting (polarized), meaning there is a larger concentration of negative ions inside the membrane than at its exterior.

When the nervous cell is stimulated (depolarized) with an impulse higher than its activation threshold (−55 mV), caused by the variation of internal concentrations of sodium (Na^+) and potassium (K^+) ions, it triggers an electrical impulse which will propagate throughout its axon with a maximum amplitude of 35 mV (Kandel et al. 2012). The stages related to variations of the action voltage within a neuron during its excitation are shown in Fig. 1.3.

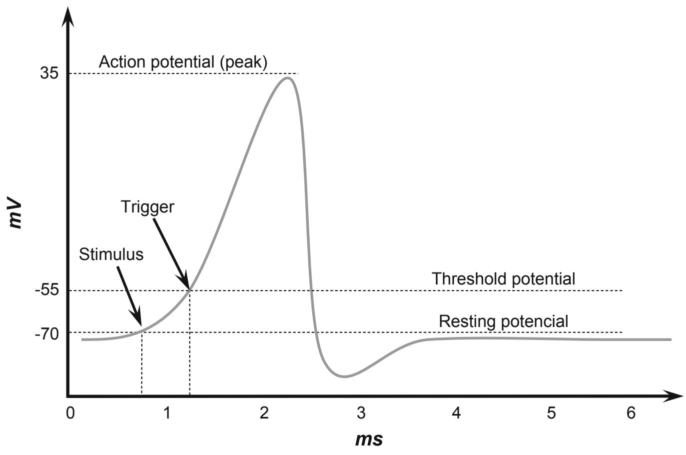


Fig. 1.3 Stages of the action potential

It is important to emphasize that the amplitude of 35 mV, the maximum value of the action voltage, is fixed and strictly satisfied for all neurons when they are stimulated, however, the signal duration in time is variable. This fact can be observed independently of the category of the neuron (connector, afferent, or efferent). As soon as the excitation process ends, the membrane will be consequently repolarized, meaning the action voltage will return to its rest voltage (−70 mV), as illustrated by Fig. 1.3.

1.3 Artificial Neuron

The artificial neural network structures were developed from known models of biological nervous systems and the human brain itself. The computational components or processing units, called artificial neurons, are simplified models of biological neurons. These models were inspired by the analysis of how a cell membrane of a neuron generates and propagates electrical impulses (Hodgkin and Huxley 1952).

The artificial neurons used in artificial neural networks are nonlinear, usually providing continuous outputs, and performing simple functions, such as gathering signals available on their inputs, assembling them according to their operational functions, and producing a response considering their innate activation functions.

The most simple neuron model that includes the main features of a biological neural network—parallelism and high connectivity—was proposed by McCulloch and Pitts (1943), and still is the most used model in different artificial neural network architectures.

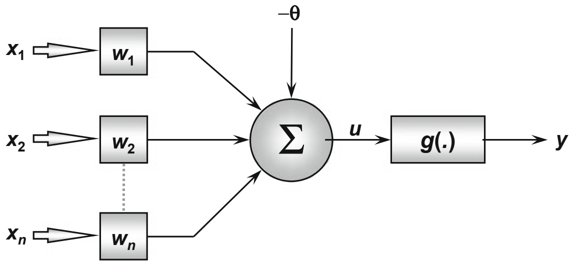


Fig. 1.4 The artificial neuron

In that model, each neuron from a network can be implemented as shown in Fig. 1.4. The multiple input signals coming from the external environment (application) are represented by the set $\{x_1, x_2, x_3, \ldots, x_n\}$, analogous to the external electrical impulses gathered by the dendrites in the biological neuron.

The weighing carried out by the synaptic junctions of the network are implemented on the artificial neuron as a set of synaptic weights $\{w_1, w_2, \ldots, w_n\}$. Analogously, the relevance of each of the $\{x_i\}$ neuron inputs is calculated by multiplying them by their corresponding synaptic weight $\{w_i\}$, thus weighting all the external information arriving to the neuron. Therefore, it is possible to verify that the output of the artificial cellular body, denoted by u, is the weighted sum of its inputs.

Considering Fig. 1.4, it is possible to see that the artificial neuron is composed of seven basic elements, namely:

(a) *Input signals* $(x_1, x_2, \ldots, x_n)$ are the signals or samples coming from the external environment and representing the values assumed by the variables of a particular application. The input signals are usually normalized in order to enhance the computational efficiency of learning algorithms.
(b) *Synaptic weights* $(w_1, w_2, \ldots, w_n)$ are the values used to weight each one of the input variables, which enables the quantification of their relevance with respect to the functionality of the neuron.
(c) *Linear aggregator* (Σ) gathers all input signals weighted by the synaptic weights to produce an activation voltage.
(d) *Activation threshold or bias* (θ) is a variable used to specify the proper threshold that the result produced by the linear aggregator should have to generate a trigger value toward the neuron output.
(e) *Activation potential* (u) is the result produced by the difference between the linear aggregator and the activation threshold. If this value is positive, i.e. if $u \geq \theta$, then the neuron produces an excitatory potential; otherwise, it will be inhibitory.
(f) *Activation function* (g) whose goal is limiting the neuron output within a reasonable range of values, assumed by its own functional image.
(g) *Output signal* (y) consists on the final value produced by the neuron given a particular set of input signals, and can also be used as input for other sequentially interconnected neurons.

The two following expressions synthesize the result produced by the artificial neuron proposed by McCulloch and Pitts:

$$u = \sum_{i=1}^{n} w_i \cdot x_i - \theta \tag{1.1}$$

$$y = g(u) \tag{1.2}$$

Thus, the artificial neuron operation can be summarized by the following steps:

(i) Present a set of values to the neuron, representing the input variables.
(ii) Multiply each input of the neuron to its corresponding synaptic weight.
(iii) Obtain the activation potential produced by the weighted sum of the input signals and subtract the activation threshold.
(iv) Applying a proper activation function to limit the neuron output.
(v) Compile the output by employing the neural activation function in the activation potential.

The activation functions can be categorized into two fundamental groups, *partially differentiable functions,* and *fully differentiable functions*, when considering their complete definition domains.

1.3.1 Partially Differentiable Activation Functions

Partially differentiable activation functions are functions with points whose first order derivatives are nonexisting. The three main functions of this category are the following: step function, bipolar step function, and symmetric ramp function.

(a) *Step function (Heaviside/Hard limiter)*
The result produced by the step function will assume unitary positive values when the neuron activation potential is greater or equal zero; otherwise, the result will be null. Thus, we have:

$$g(u) = \begin{cases} 1, & \text{if } u \geq 0 \\ 0, & \text{if } u < 0 \end{cases} \tag{1.3}$$

The step function graphical representation is illustrated in Fig. 1.5.

(b) *Bipolar step function or Signal function (Symmetric hard limiter)*
The result produced by this function will assume unitary positive values when the neuron activation potential is greater than zero; null value when the potential is also null; and negative unitary values when the potential is less than zero. This behavior in mathematical notation is:

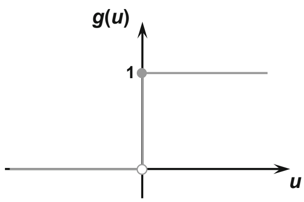


Fig. 1.5 The step activation function

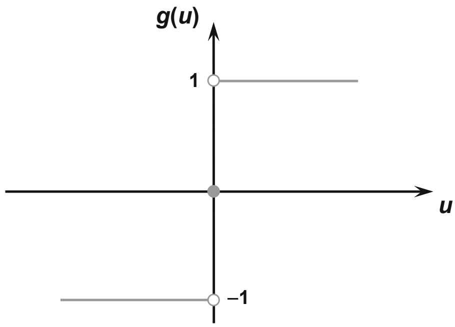


Fig. 1.6 The bipolar step activation function

$$g(u) = \begin{cases} 1, & \text{if } u > 0 \\ 0, & \text{if } u = 0 \\ -1, & \text{if } u < 0 \end{cases} \tag{1.4}$$

The graphical representation of this function is illustrated in Fig. 1.6.
In problems involving pattern classification, the bipolar step function can be approximated by the following expression:

$$g(u) = \begin{cases} 1, & \text{if } u \geq 0 \\ -1, & \text{if } u < 0 \end{cases} \tag{1.5}$$

In this circumstance, another alternative is to maintain the neuron output unchanged, thus:

$$g(u) = \begin{cases} 1, & \text{if } u > 0 \\ \text{previous output}, & \text{if } u = 0 \\ -1, & \text{if } u < 0 \end{cases} \tag{1.6}$$

(c) *Symmetric ramp function*
The values returned by this function are equal to the values of the activation potential themselves when defined within the range $[-a, a]$, and limited to the limit values otherwise. The mathematical notation for this behavior is as follows:

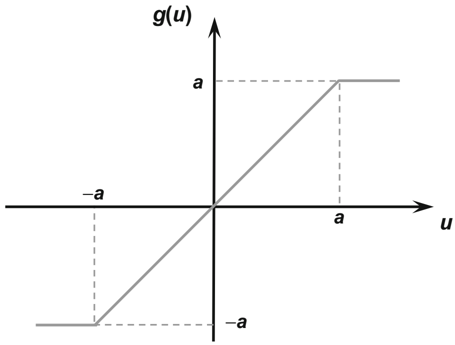


Fig. 1.7 The symmetric ramp activation function

$$g(u) = \begin{cases} a, & \text{if } u > a \\ u, & \text{if } -a \leq u \leq a \\ -a, & \text{if } u < a \end{cases} \tag{1.7}$$

The graphical representation of this function is illustrated in Fig. 1.7.

1.3.2 Fully Differentiable Activation Functions

Fully differentiable activation functions are those whose first order derivatives exist for all points of their definition domain. The four main functions of this category, which can be employed on artificial neural networks, are the logistic function, hyperbolic tangent, Gaussian function and linear function.

(a) *Logistic function*
The output result produced by the logistic function will always assume real values between zero and one. Its mathematical expression is given by:

$$g(u) = \frac{1}{1 + e^{-\beta \cdot u}}, \tag{1.8}$$

where β is a real constant associated with the function slope in its inflection point. Figure 1.8 illustrates the behavior of this function.
Figure 1.9 shows the behavior of the logistic function when the slope parameter β changes.
From the analysis of Fig. 1.9, it is possible to conclude that the geometric format of the logistic activation function is similar to that of the step function, when β is very high, i.e., tending to infinity. However, in contrast to the step function, the logistic function is fully differentiable in its entire definition domain.

Fig. 1.8 The logistic activation function

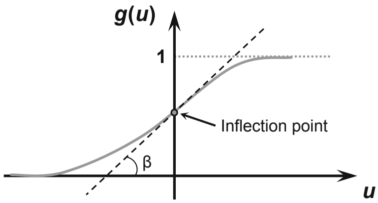
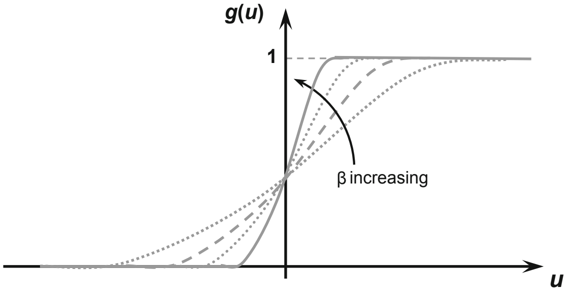


Fig. 1.9 Influence of the parameter β in the logistic activation function

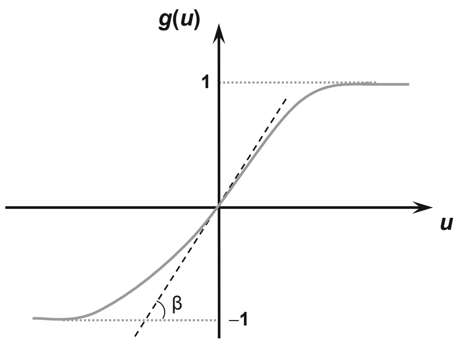


(b) *Hyperbolic tangent function*
The output result, unlike the case of the logistic function, will always assume real values between −1 and 1, with the following mathematical expression:

$$g(u) = \frac{1 - e^{-\beta \cdot u}}{1 + e^{-\beta \cdot u}} \ , \tag{1.9}$$

where β is also associated with the slope of the hyperbolic tangent function in its inflection point. The graphical representation of this function is illustrated by Fig. 1.10.
Figure 1.11 also illustrates the behavior of the hyperbolic tangent function when the parameter β changes.

Fig. 1.10 The hyperbolic tangent activation function



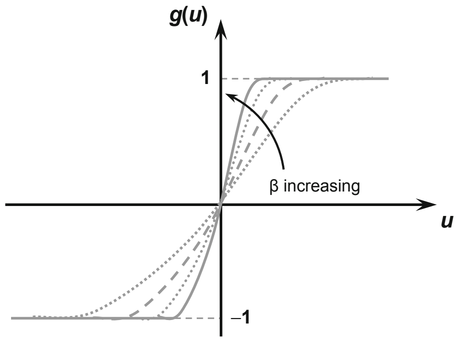


Fig. 1.11 Influence of the β parameter on the hyperbolic tangent activation function

As observed in Fig. 1.11, the higher the value of β, the higher the slope the hyperbolic tangent function will have—as in the case of the logistic function—and it will approximate to the bipolar step function (signal) when the β value is very high.

It is important to note that both logistic and hyperbolic tangent functions belong to a family of functions called sigmoidal.

(c) *Gaussian function*

In the case of Gaussian activation functions, the neuron output will produce equal results for those activation potential values $\{u\}$ placed at the same distance from its center (average). The curve is symmetric to this center and the Gaussian function is given by:

$$g(u) = e^{-\frac{(u-c)^2}{2\sigma^2}}, \tag{1.10}$$

where c is the parameter that defines the center of the Gaussian function and σ denotes the associated standard deviation, that is, how scattered (dispersed) is the curve in relation to its center. The graphical representation of this function is illustrated by Fig. 1.12.

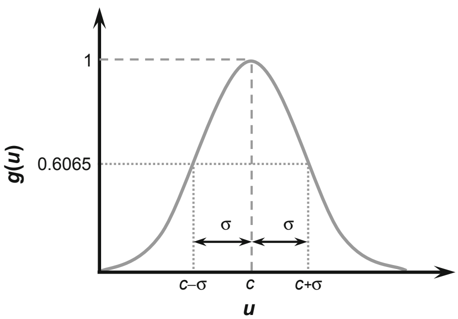
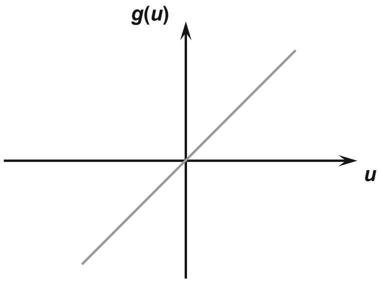


Fig. 1.12 The Gaussian activation function

Fig. 1.13 The linear activation function



It is possible to observe on this figure that the standard deviation parameter $\{\sigma\}$ is directly associated with the inflection points of the Gaussian function, with σ^2 indicating its variance.

(d) *Linear function*

The linear activation function, or identity function, produces output results equal to the activation potential $\{u\}$, having its mathematical expression given by:

$$g(u) = u \tag{1.11}$$

The graphical representation of this function is illustrated in Fig. 1.13.

One application of the linear activation functions is in artificial neural networks performing universal curve fitting (function approximation), to map the behavior of the input/output variables of a particular process, as it is discussed in Sect. 5.4.

1.4 Performance Parameters

To relate the operation of both artificial and biological neurons, Table 1.2 presents a comparison between their features of performance.

It is possible to observe that the processing time of artificial neurons is lower than that of biological neurons. On the other hand, the cerebral processing is countlessly faster, in most cases, than any artificial neural network, since neurons from biological neural networks operate with high degree of parallelism. Neurons from artificial neural networks have very limited parallelism capabilities, because most computers are built with sequential machines (Haykin 2009; Faggin 1991).

Table 1.2 Comparative chart between artificial and biological neurons

Parameter	Artificial neuron	Biological neuron
Energetic efficiency (operation/second)	10^{-6} J	10^{-16} J
Processing time (operation/neuron)	10^{-9} s (clock on the order GHz)	10^{-3} s
Processing mechanism	Usually sequential	Usually parallel

The speed parameter of artificial neural networks is essentially related to the number of operations per second performed by computers. Considering a clock on the order of gigahertz, the processing period of artificial neurons are in the magnitude of nanoseconds.

1.5 Exercises

1. Explain how an artificial neuron operates.
2. Describe what are the main goals of activation functions.
3. Make an analogy between the elements composing artificial and biological neurons.
4. Write about the importance of the activation threshold (or bias).
5. Thinking about the features of artificial neural networks, explain what is learning from experience and generalization capability.
6. Write about the main mathematical features which can be verified on both the logistic and hyperbolic tangent activation functions.
7. Find the analytical expressions of the first order derivatives of the logistic and hyperbolic tangent.
8. For a particular problem, it is possible to use the logistic or hyperbolic functions as the activation function. Regarding hardware implementation, write about the eventual features to be considered for selecting one of them.
9. Given that individual operation on artificial neurons are executed faster when compared to biological neurons, explain why many tasks performed by the human brain produce results faster than a microcomputer.
10. What are the main categories of problems which can be addressed by artificial neural networks?