Introduction to Artificial Neural Networks Training and Applications

M.N. Vrahatis[1], G.D. Magoulas[2],
K.E. Parsopoulos[1], V.P. Plagianakos[1]

vrahatis@math.upatras.gr, George.Magoulas@brunel.ac.uk, kostasp@math.upatras.gr, vpp@math.upatras.gr

[1]*Department of Mathematics, University of Patras, GR-261 10, Patras.*
[2]*Department of Information Systems and Computing, Brunel University, UB8 3PH, U.K.*

BACKGROUND

Scientific interest in models of neuronal networks or artificial neural networks mainly arises from their potential ability to perform interesting computational tasks. Nodes, or artificial neurons, in neuronal network models are usually considered as simplified models of biological neurons, i.e. real nerve cells, and the connection weights between nodes resemble to synapses between neurons. In fact, artificial neurons are much simpler than biological neurons. But, for the time being, it is far from clear how much of this simplicity is justified because, as yet, we have only poor understanding of neuronal functions in complex biological networks. Artificial neural networks (ANNs) provide to computing an alternative algorithmic model, which is biologically motivated: the computation is massively distributed and parallel and the learning replaces a priori program development, i.e. ANNs develop their functionality based on training (sampled) data.

In neural network learning the objective is usually to minimize a cost function defined as the multi-variable error function of the network. This perspective gives some advantage to the development of effective learning algorithms, because the problem of minimizing a function is well known in the field of numerical analysis.

However, due to the special characteristics of the neural networks, learning algorithms can be trapped in an undesired local minimum of the error function, since most of them are based on local search methods and have no mechanism that allows them to escape the influence of such undesired local minima.

ANNs have been applied to an increasing number of world problems of considerable complexity. Their most important advantage is in solving problems that are too complex for conventional technologies - problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be found. In general, because of their abstraction from the biological brain, ANNs are well suited to problems that people are good at solving but computers are not. These problems include pattern recognition and forecasting (which requires the recognition of trends in data). However, unlike the human capability in pattern recognition, the ANN's capability is not affected by factors such as fatigue, working conditions, emotional state, and compensation.

The ANN approach to medical information processing has several benefits.

- Training by examples instead of rules.
- It is automated.

- It eliminates issues associated with human fatigue and habituation.
- It enables rapid identification.
- It enables analysis of conditions and diagnosis in real time.

The ANN approach to the analysis of data will see extensive application to biomedical problems in the next few years. It has already been successfully applied to various areas of medicine, such as diagnostic aides, biochemical analysis, image analysis, and drug development.

MODELS AND BASIC ARCHITECURES

Artificial Neural Networks (ANNs) are nonparametric regression models. They can capture any phenomena, to any degree of accuracy (depending on the adequacy of the data and the power of the predictors), without prior knowledge of the phenomena. Further, ANNs can be represented, not only as formulae, but also as graphical models. Graphical models can improve the manipulation, and understanding, of statistical structures. ANNs are a powerful method for capturing complex phenomena, but their use requires a paradigm shift, from exploratory analysis of the data to exploratory analysis of the model.

Each ANN is consisted of information-processing units called artificial neurons which are simplified models of the biological neurons. Figures 1 and 2 show a biological and an artificial model. The artificial neuronal model consists of three basic elements:

- A set of *synapses* or *connecting links*, each of which is characterized by a *weight*. Unlike a synapse in the brain, the synaptic weight of an artificial neuron may lie in a range that includes negative as well as positive values.

- An *adder* for summing the input signals, weighted by the respective synapses of the neuron.
- An *activation function* for limiting the amplitude of the output of a neuron.

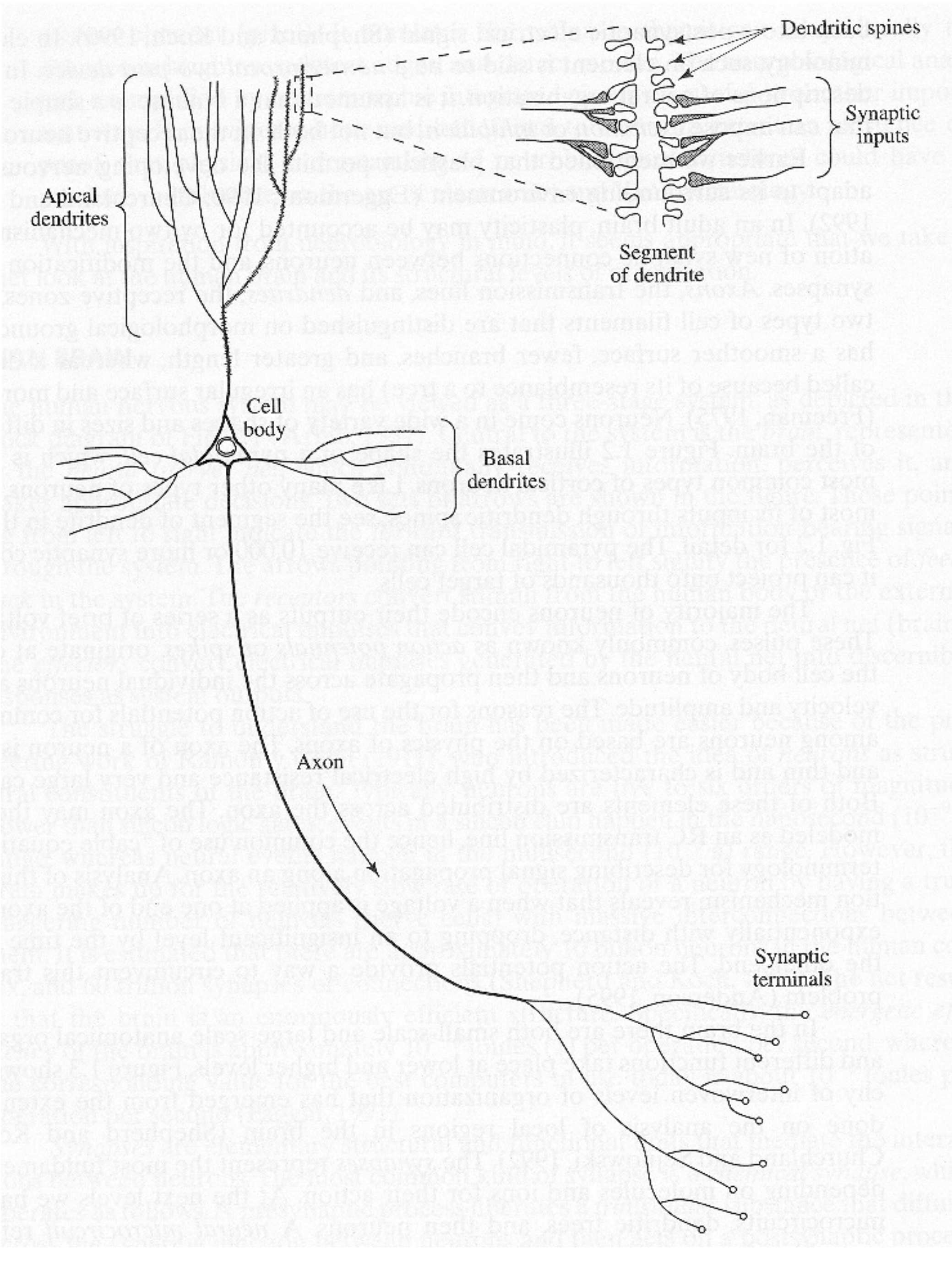

Figure 1: The Biological Neuron

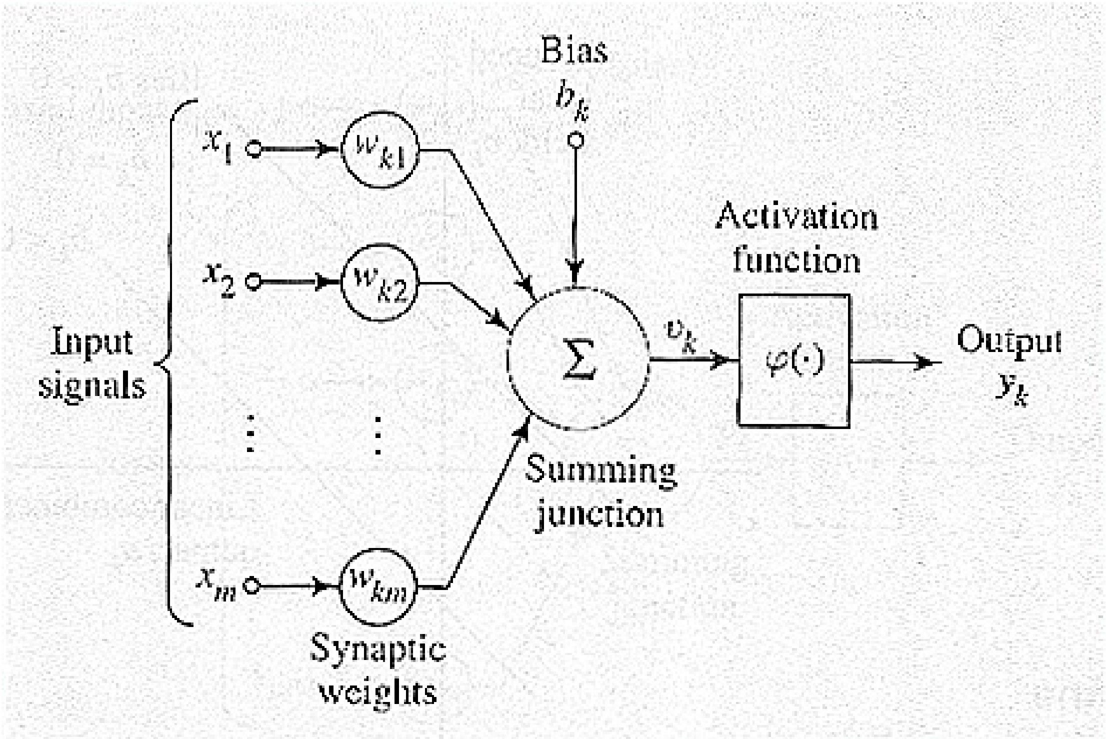

Figure 2: The Artificial Neuron's model

In principle, ANNs can compute any computable function, i.e., they can do everything a normal digital computer can do, or perhaps even more, under some assumptions of doubtful practicality. In practice, ANNs are especially useful for classification and function approximation/mapping problems which are tolerant of some imprecision and have lots of training data available, but to which hard and fast rules (such as those that might be used in an expert system) cannot easily be applied.

Almost any finite-dimensional vector function on a compact set can be approximated to arbitrary precision by *Feedforward* ANNs if you have enough data and enough computing resources.

Feedforward ANNs have their neurons organized in several layers and the output of a neuron can be transmitted only to neurons of the next layer. An example of such a network with one hidden layer can be seen in Figure 3.

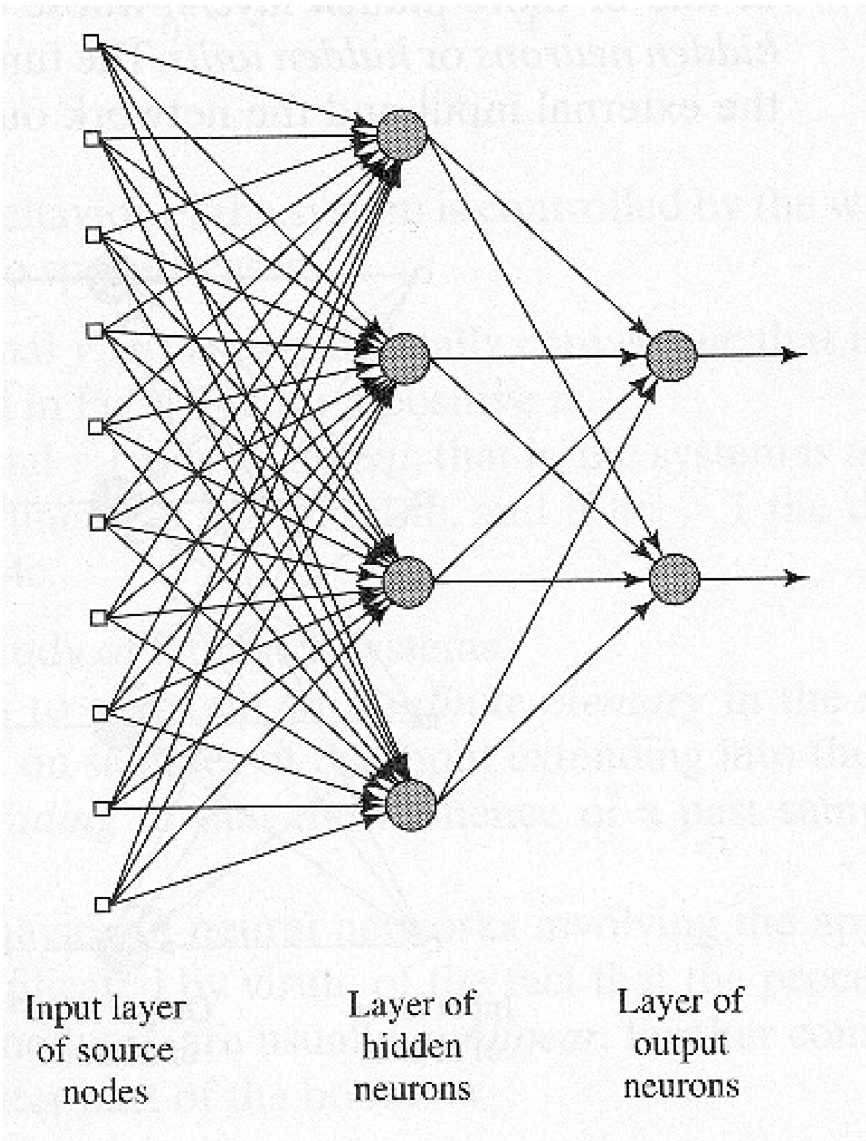

Figure 3: A Feedforward ANN with one Hidden Layer

Feedforward networks with a single hidden layer, using threshold or sigmoid activation functions, are universally consistent estimators of binary classifications under some assumptions. Note that these results are stronger than the universal approximation theorems that merely show the existence of weights for arbitrarily accurate approximations, without demonstrating that such weights can be obtained by learning.

Unfortunately, the above consistency results depend on one impractical assumption: that the networks are trained by an error minimization technique that comes arbitrarily close to the global minimum. Such minimization is computationally intractable except for small or simple problems. In practice, however, you can usually get good results without doing a global optimization; instead, one can use multiple random weight initializations. Finally, it is important to realize that there are no methods for training ANNs that can magically create information that is not contained in the training data.

ANNs have many applications to the analysis of data in biomedical problems as will be discussed in the next paragraph.

MEDICAL DATA PREPROCESSING

Medical data typically requires a large amount of preprocessing in order to be useful. There is often *redundancy in the data*; for example age may appear in several places. *Erroneous data* are very common. Finally, medical data are frequently *sparse*; when a structure is imposed on medical data much of the structure remains empty for a large portion of the population due to the breadth required of any structure.

So, a robust data preprocessing system is required in order to draw any kind of knowledge from even medium sized medical data sets. The data must not only be *cleaned of errors and redundancy*, but organized in a fashion which makes sense for the problem, i.e. the data must be organized so that the benefits of using neural networks are maximized.

A MEDICAL IMAGING APPLICATION

An ANN architecture with 16 input neurons, 30 hidden neurons and 2 output neurons has been used for detecting two different types of abnormalities in colonoscopic images taken from two different colons. Image 1 (Figure 4) is macroscopically a Type-IIIs lesion according to Kudo (1996). Histologically it is a *low grade cancer*. Image 2 (Figure 2) is macroscopically a Type-V lesion according to Kudo (1996). Histologically it is a *moderately differentiated carcinoma*. Textures from 10 normal and 10 abnormal tissue samples have been randomly chosen from each image and used for training the MLP. The performance of the trained MLP has been tested on a new set of 80 texture samples (40 normal and 40 abnormal) randomly obtained from the two images.

MODELING COGNITION

Much work is currently under development for simulating human consciousness and emotion. Consciousness is still one of the world's great mysteries. Artificial NNs may be useful for modeling some aspects of, or prerequisites for consciousness, such as perception and cognition, but ANNs provide no insight so far into what Chalmers (1996) calls the “hard problem”:

> *“Many books and articles on consciousness have appeared in the past few years, and one might think we are making progress. But on a closer look, most of this work leaves the hardest problems about consciousness untouched. Often, such work addresses what might be called the "easy problems" of consciousness: How does the brain process environmental stimulation? How does it integrate information? How do we produce reports on internal states? These are important questions, but to answer them is not to solve the hard problem: Why is all this processing accompanied by an experienced inner life?”*

REFERENCES

David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York and Oxford, 1996.

Shin-ei Kudo, *Early Colorectal Cancer: Detection of Depressed Types Colorectal Carcinomas*, Lippincott Williams & Wilkins, Philadelphia, U.S.A., 1996.