Physica A 185 (1992) 402–410
North-Holland

PHYSICA A

Learning in feed-forward neural networks by improving the performance

Mirta B. Gordon[a,1], Pierre Peretto[b] and
Miguel Rodriguez-Girones[b]
[a]*SPSMS/MDN, DRFMC, Centre d'Etudes Nucléaires de Grenoble, 85 X, 38041 Grenoble Cedex, France*
[b]*SP2M/PSC, DRFMC, Centre d'Etudes Nucléaires de Grenoble, 85 X, 38041 Grenoble Cedex, France*

Statistical mechanics is used to derive a new learning rule for a feed-forward neural network with one hidden layer. Generalization to multilayer neural networks is straightforward, and proceeds in the same way as backpropagation.

We consider a neural network as a physical system, that can be in different states. There are as many possible states as patterns in the learning set. The energy of each level is proportional to the stability of the corresponding pattern. The statistical mechanics free energy of the system, which we call *performance*, is a maximum for the synaptic strengths that stabilize all the patterns.

We propose a learning algorithm that looks for synaptic strengths that maximize the network's performance. Patterns with lower stabilities are more effective in driving the learning process because they have a higher statistical weight. Taking different temperatures for different layers improves the results.

1. Introduction

Learning and generalizing with neural networks is a promising way of tackling problems which do not have any known algorithm that solves them. Much work has been done in recent years to understand the general properties of neural networks, and statistical mechanics proved to be a powerful tool in this field [1, 2].

Learning in neural networks is the process by which the connexions between neurons, the synaptic efficacies, are modified in order that a given set of input–output relations between patterns are reproduced. Most of our present knowledge concerns networks containing only visible neurons. However, some problems require the introduction of hidden units. A typical example is the exclusive or – the XOR – of two binary inputs. The minimal feed-forward

[1] Also with the Centre National de la Recherche Scientifique (CNRS).