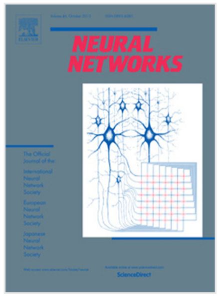
Accepted Manuscript



A theory of local learning, the learning channel, and the optimality of backpropagation

Pierre Baldi, Peter Sadowski

PII: S0893-6080(16)30088-0
DOI: http://dx.doi.org/10.1016/j.neunet.2016.07.006
Reference: NN 3641

To appear in: *Neural Networks*

Received date: 26 October 2015
Revised date: 31 May 2016
Accepted date: 13 July 2016

A Theory of Local Learning, the Learning Channel, and the Optimality of Backpropagation

PIERRE BALDI* and PETER SADOWSKI
Department of Computer Science
University of California, Irvine
Irvine, CA 92697-3435

pfbaldi,pjsadows@uci.edu

Abstract

In a physical neural system, where storage and processing are intimately intertwined, the rules for adjusting the synaptic weights can only depend on variables that are available locally, such as the activity of the pre- and post-synaptic neurons, resulting in *local learning rules*. A systematic framework for studying the space of local learning rules is obtained by first specifying the nature of the local variables, and then the functional form that ties them together into each learning rule. Such a framework enables also the systematic discovery of new learning rules and exploration of relationships between learning rules and group symmetries. We study polynomial local learning rules stratified by their degree and analyze their behavior and capabilities in both linear and non-linear units and networks. Stacking local learning rules in deep feedforward networks leads to *deep local learning*. While deep local learning can learn interesting representations, it cannot learn complex input-output functions, even when targets are available for the top layer. Learning complex input-output functions requires *local deep learning* where target information is propagated to the deep layers through a backward learning channel. The nature of the propagated information about the targets and the structure of the learning channel partition the space of learning algorithms. For any learning algorithm, the *capacity* of the learning channel can be defined as the number of bits provided about the error gradient per weight, divided by the number of required operations per weight. We estimate the capacity associated with several learning algorithms and show that backpropagation far outperforms them by simultaneously maximizing the information rate and minimizing the computational cost. This result is shown also to be true for recurrent networks, by unfolding them in time. The theory clarifies the concept of Hebbian learning, establishes the power and limitations of local learning rules, provides a formal analysis of the optimality of backpropagation, and explains the sparsity of the space of learning rules discovered so far.

Keywords: machine learning; neural networks; recurrent networks; recursive networks; supervised learning; unsupervised learning; deep learning; backpropagation; learning rules; Hebbian learning.

1 Introduction

The deep learning problem can be viewed as the problem of learning the connection weights of a large computational graphs, in particular the weights of the deep connections that are far away from the inputs or outputs [51]. In spite of decades of research, only very few algorithms have been proposed to try to address this task. Among the most important ones, and somewhat in opposition to each other, are backpropagation [50] and Hebbian learning [26]. Backpropagation has been the dominant algorithm, at least in terms of successful applications, which have ranged over the years from computer vision [35] to high-energy physics [10]. In spite of many attempts, no better algorithm has been found, at least within the standard supervised learning framework. In contrast to backpropagation which is a well defined algorithm–stochastic gradient descent–Hebbian learning has remained a more nebulous concept, often associated with notions of biological and unsupervised learning. While less successful than backpropagation in applications, it has periodically inspired the development of theories aimed at capturing the essence of neural learning [26, 22, 29]. Within this general context, the goal of this work is to create a precise framework to organize and study the space of learning rules and their properties and address several questions, in particular: (1) What is Hebbian learning? (2) What are the capabilities and limitations of Hebbian learning? (3) What are the connections between Hebbian learning and backpropagation? (4) Are there other learning algorithms better than backpropagation? These questions are addressed in two parts: the first part focuses on Hebbian learning, the second part on backpropagation.

*Contact author