Statistical Mechanics of Learning

Statistical Mechanics of Learning

A. Engel
Otto von Guericke University, Magdeburg, Germany

C. Van den Broeck
Limburgs Universitair Centrum, Diepenbeek, Belgium

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/14pt. *System* LaTeX 2_ε [DBD]

A catalogue record of this book is available from the British Library

ISBN 0 521 77307 5 hardback
ISBN 0 521 77479 9 paperback

Contents

Preface

Understanding intelligent behaviour has always been fascinating to both laymen and scientists. The question has become very topical through the concurrence of a number of different issues. First, there is a growing awareness of the computational limits of serial computers, while parallel computation is gaining ground, both technically and conceptually. Second, several new non-invasive scanning techniques allow the human brain to be studied from its collective behaviour down to the activity of single neurons. Third, the increased automatization of our society leads to an increased need for algorithms that control complex machines performing complex tasks. Finally, conceptual advances in physics, such as scaling, fractals, bifurcation theory and chaos, have widened its horizon and stimulate the modelling and study of complex non-linear systems. At the crossroads of these developments, *artificial neural networks* have something to offer to each of them.

The observation that these networks can learn from examples and are able to discern an underlying rule has spurred a decade of intense theoretical activity in the statistical mechanics community on the subject. Indeed, the ability to infer a rule from a set of examples is widely regarded as a sign of intelligence. Without embarking on a thorny discussion about the nature or definition of intelligence, we just note that quite a few of the problems posed in standard IQ tests are exactly of this nature: given a sequence of objects (letters, pictures, ...) one is asked to continue the sequence “meaningfully”, which requires one to decipher the underlying rule. We can thus hope that a theoretical study of learning from examples in simple, well understood scenarios will shed some light on how intelligent behaviour emerges or operates.

Artificial systems which can learn from examples also offer a very appealing alternative to rule and program-driven systems in cases where explicit algorithms are difficult to find. After all, most of the things we do have been learned from experience and are hard to formulate in terms of rules. A simple example is the sorting of passport photographs according to whether the person shown is male or female. Every human more than five years old performs extraordinarily well on this task, with respect to both error rate and speed. Yet no present-day computer can handle this problem. This is *not* because of its limited speed or power, but due to the inability of the programmer to convey to the machine what to do. It is an extremely attractive idea in problems as complex as the above pattern recognition task to instruct the computer by means of a learning session using examples.

Furthermore, machines which learn from examples typically operate in a parallel and distributed fashion, and therefore share several fundamental characteristics with the human brain, such as fault tolerance, flexibility, and high computational power. With the promise of parallel computers tackling the hardest computational problems, one may hope to bridge the increasing gap between hardware and software by studying how to construct and train machines which learn from examples. At the same time, detailed neural networks have been constructed that mimic quite well the functioning of their biological counterparts. These biologically inspired networks together with their fully artificial counterparts are increasingly being used in a variety of practical applications, from robotics through speech and image processing to control and decision making.

The topic under consideration is in fact old and carries a distinct interdisciplinary flavour. The theory of learning is a classic subject in psychology, philosophy, and pedagogy. But it is also being addressed in other scientific disciplines. Computational learning theory explores the theoretical boundaries of how well simple devices like the Turing machine or more abstract constructions, such as classifier systems, can learn. Mathematical statistics and information theory tell us how much we can infer from observed data. Even though neural networks have become an active subfield of research in statistical mechanics, extending well beyond the incipient connection between associative memories and spin glasses, it is a newcomer in the field compared to the classical disciplines involved. Correspondingly it is only able to deal with rather simple learning scenarios so far. Nevertheless, starting from the new central dogma that information processing in artificial neural networks is a *collective* property well beyond the ability of the individual constituents, and borrowing from an impressive arsenal of both conceptual and technical tools developed over the last hundred years, it is beginning to contribute stimulating new results from a fresh point of view.

The present book is about the statistical mechanics of learning, providing an introduction both to the basic notions and to the relevant techniques used to obtain quantitative results. Care has been taken to make most of the concepts accessible without following all calculations in detail. On the other hand, being a part of theoretical physics, the field relies on mathematical modelling and thorough understanding will be impossible without having control of the relevant techniques.

A major part of the book deals with the perceptron, which is the basic building block for neural networks. While the computational capabilities of the perceptron are very limited, all its properties can be discussed in full analytic detail. After elucidating various aspects of perceptron learning, including a detailed description of the basic setup, a review of the most pertinent learning rules, the storage problem, and discontinuous, unsupervised, and on-line learning in chapters 1–11, we discuss multilayer networks in chapters 12 and 13. These chapters are supplemented by problems forming an integral part of the material. In the discussion section at the end of every chapter, we include a brief review of the literature. These references are cited for guidance and reflect our own interest and background. We apologize to authors whose work is not cited adequately. In the final chapter we glance at some related problems from the statistical mechanics of complex systems such as support vector machines, computationally hard problems, error-correcting codes, and game theory. Most of the more technical aspects have been collected in the appendices, where the basic computational techniques are presented in great detail.

We would like to take the opportunity to express our gratitude to the many friends and colleagues together with whom our understanding of the field has been shaped, notably to Michael Biehl, Marc Bouten, Nestor Caticha, Mirta Gordon, Geza Györgyi, John Hertz, Ido Kanter, Wolfgang Kinzel, Reimar Kühn, Rémi Monasson, Manfred Opper, Juan Manuel Parrondo, Peter Reimann, Lüder Reimers, Pál Rujan, David Saad, Holm Schwarze, Sara Solla, Timothy Watkin, Michael Wong, Richardo Zecchina, Annette Zippelius as well as to our students Ioana Bena, Johannes Berg, Geert Jan Bex, Mauro Copelli, Wolfgang Fink, Eddy Lootens, Peter Majer, Doerthe Malzahn, Jan Schietse and Martin Weigt. We regret not having had the privilege of meeting Elizabeth Gardner, whose groundbreaking contributions will continue to inspire the whole field of neural networks.

We are indebted to Alexander Schinner for helpful assistance in all kinds of computer problems. We would also like to thank Michael Biehl, Mauro Copelli, Stephan Mertens, and Richard Metzler for reading preliminary versions of parts of this book and making useful comments. Special thanks are due to Johannes Berg, who literally read every line of the manuscript (except for these), pointed out mistakes, tracked down inconsistencies, removed colons, improved arguments, corrected misspellings, and suggested references.

Part of the work related to this book was performed during a stay in 1999 at the Max-Planck-Institut für komplexe Systeme in Dresden, whom we would like to thank for hospitality and excellent working conditions.