Springer Series in Statistics

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

László Györfi Michael Kohler
Adam Krzyżak Harro Walk

A Distribution-Free Theory of Nonparametric Regression

With 86 Figures

Springer

László Györfi
Department of Computer Science and Information Theory
Budapest University of Technology and Economics
1521 Stoczek, U.2.
Budapest
Hungary
gyorfi@inf.bme.hu

Michael Kohler
Fachbereich Mathematik
Universität Stuttgart
Pfaffenwaldring 57
70569 Stuttgart
Germany
kohler@mathematik.uni-stuttgart.de

Adam Krzyżak
Department of Computer Science
Concordia University
1455 De Maisonneuve Boulevard West
Montreal, Quebec, H3G 1M8
Canada
krzyzak@cs.concordia.ca

Harro Walk
Fachbereich Mathematik
Universität Stuttgart
Pfaffenwaldring 57
70569 Stuttgart
Germany
walk@mathematik.uni-stuttgart.de

Library of Congress Cataloging-in-Publication Data
A distribution-free theory of nonparametric regression / László Györfi . . . [et al.].
p. cm. — (Springer series in statistics)
Includes bibliographical references and index.
ISBN 0-387-95441-4 (alk. paper)
1. Regression analysis. 2. Nonparametric statistics. 3. Distribution (Probability theory)
I. Györfi, László. II. Series.
QA278.2 .D57 2002
519.5′36—dc21 2002021151

ISBN 0-387-95441-4 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 10866288

Typesetting: Pages created by the authors using a Springer TEX macro package.

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

To our families:

Kati, Kati, and Jancsi
Judith, Iris, and Julius
Henryka, Jakub, and Tomasz
Hildegard

Preface

The *regression estimation problem* has a long history. Already in 1632 Galileo Galilei used a procedure which can be interpreted as fitting a linear relationship to contaminated observed data. Such fitting of a line through a cloud of points is the classical linear regression problem. A solution of this problem is provided by the famous principle of least squares, which was discovered independently by A. M. Legendre and C. F. Gauss and published in 1805 and 1809, respectively. The principle of least squares can also be applied to construct *nonparametric* regression estimates, where one does not restrict the class of possible relationships, and will be one of the approaches studied in this book.

Linear regression analysis, based on the concept of a regression function, was introduced by F. Galton in 1889, while a probabilistic approach in the context of multivariate normal distributions was already given by A. Bravais in 1846. The first nonparametric regression estimate of local averaging type was proposed by J. W. Tukey in 1947. The partitioning regression estimate he introduced, by analogy to the classical partitioning (histogram) density estimate, can be regarded as a special least squares estimate.

Some aspects of nonparametric estimation had already appeared in belletristic literature in 1930/31 in *The Man Without Qualities* by Robert Musil (1880-1942) where, in Section 103 (first book), methods of partitioning estimation are described: "... as happens so often in life, you ... find yourself facing a phenomenon about which you can't quite tell whether it is a law or pure chance; that's where things acquire a human interest. Then you translate a series of observations into a series of figures, which you divide into categories to see which numbers lie between this value and that,

and the next, and so on .... You then calculate the degree of aberration, the mean deviation, the degree of deviation from some arbitrary value ... the average value ... and so forth, and with the help of all these concepts you study your given phenomenon" (cited from page 531 of the English translation, Alfred A. Knopf Inc., Picador, 1995).

Besides its long history, the problem of regression estimation is of increasing importance today. Stimulated by the tremendous growth of information technology in the past 20 years, there is a growing demand for procedures capable of automatically extracting useful information from massive highly-dimensional databases that companies gather about their customers. One of the fundamental approaches for dealing with this "data-mining problem" is regression estimation. Usually there is little or no *a priori* information about the data, leaving the researcher with no other choice but a nonparametric approach.

This book presents a modern approach to nonparametric regression with random design. The starting point is a prediction problem where minimization of the mean squared error (or L_2 risk) leads to the regression function. If the goal is to construct an estimate of this function which has mean squared prediction error close to the minimum mean squared error, then this goal naturally leads to the L_2 error criterion used throughout this book.

We study almost all known regression estimates, such as classical local averaging estimates including kernel, partitioning, and nearest neighbor estimates, least squares estimates using splines, neural networks and radial basis function networks, penalized least squares estimates, local polynomial kernel estimates, and orthogonal series estimates. The emphasis is on the *distribution-free* properties of the estimates, and thus most consistency results presented in this book are valid for all distributions of the data. When it is impossible to derive distribution-free results, as is the case for rates of convergence, the emphasis is on results which require as few constraints on distributions as possible, on distribution-free inequalities, and on adaptation.

Our aim in writing this book was to produce a self-contained text intended for a wide audience, including graduate students in statistics, mathematics, computer science, and engineering, as well as researchers in these fields. We start off with elementary techniques and gradually introduce more difficult concepts as we move along. Chapters 1–6 require only a basic knowledge of probability. In Chapters 7 and 8 we use exponential inequalities for the sum of independent random variables and for the sum of martingale differences. These inequalities are proven in Appendix A. The remaining part of the book contains somewhat more advanced concepts, such as almost sure convergence together with the real analysis techniques given in Appendix A. The foundations of the least squares and penalized least squares estimates are given in Chapters 9 and 19, respectively.

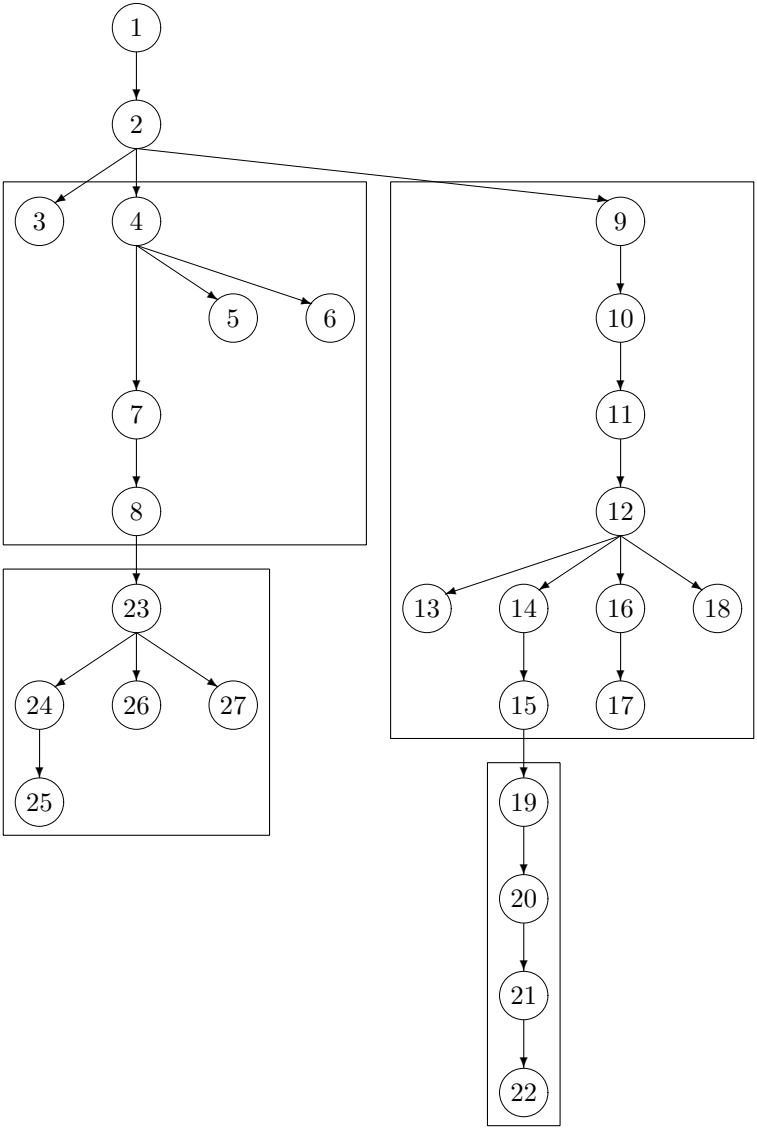

Figure 1. The structure of the book.

The structure of the book is shown in Figure 1. This figure is a precedence tree which could assist an instructor in organizing a course based on this

book. It shows the sequence of chapters needed to be covered in order to understand a particular chapter. The focus of the chapters in the upper-left box is on local averaging estimates, in the lower-left box on strong consistency results, in the upper-right box on least squares estimation, and in the lower-right box on penalized least squares.

We would like to acknowledge the contribution of many people who influenced the writing of this book. Luc Devroye, Gábor Lugosi, Eric Regener, and Alexandre Tsybakov made many invaluable suggestions leading to conceptual improvements and better presentation. A number of colleagues and friends have, often without realizing it, contributed to our understanding of nonparametrics. In particular we would like to thank in this respect Paul Algoet, Andrew Barron, Peter Bartlett, Lucien Birgé, Jan Beirlant, Alain Berlinet, Sándor Csibi, Miguel Delgado, Jürgen Dippon, Jerome Friedman, Wlodzimierz Greblicki, Iain Johnstone, Jack Koplowitz, Tamás Linder, Andrew Nobel, Mirek Pawlak, Ewaryst Rafajlowicz, Igor Vajda, Sara van de Geer, Edward van der Meulen, and Sid Yakowitz. András Antos, András György, Michael Hamers, Kinga Máthé, Dániel Nagy, Márta Pintér, Dominik Schäfer and Stefan Winter provided long lists of mistakes and typographical errors. Sándor Győri drew the figures and gave us advice and help on many LaTeX-problems. John Kimmel was helpful, patient and supportive at every stage.

In addition, we gratefully acknowledge the research support of the Budapest University of Technology and Economics, the Hungarian Academy of Sciences (MTA SZTAKI, AKP, and MTA IEKCS), the Hungarian Ministry of Education (FKFP and MÖB), the University of Stuttgart, Deutsche Forschungsgemeinschaft, Stiftung Volkswagenwerk, Deutscher Akademischer Austauschdienst, Alexander von Humboldt Stiftung, Concordia University, Montreal, NSERC Canada, and FCAR Quebec.

Early versions of this text were tried out at a DMV seminar in Oberwolfach, Germany, and in various classes at the Carlos III University of Madrid, the University of Stuttgart, and at the International Centre for Mechanical Sciences in Udine. We would like to thank the students there for useful feedback which improved this book.

László Györfi, Budapest, Hungary
Michael Kohler, Stuttgart, Germany
Adam Krzyżak, Montreal, Canada
Harro Walk, Stuttgart, Germany

June 6, 2002

Contents