CAMBRIDGE MONOGRAPHS ON APPLIED AND COMPUTATIONAL MATHEMATICS

25 **Algebraic Geometry and Statistical Learning Theory**

The *Cambridge Monographs on Applied and Computational Mathematics* reflect the crucial role of mathematical and computational techniques in contemporary science. The series publishes expositions on all aspects of applicable and numerical mathematics, with an emphasis on new developments in this fast-moving area of research.

State-of-the-art methods and algorithms as well as modern mathematical descriptions of physical and mechanical ideas are presented in a manner suited to graduate research students and professionals alike. Sound pedagogical presentation is a prerequisite. It is intended that books in the series will serve to inform a new generation of researchers.

The series includes titles in the *Library of Computational Mathematics*, published under the auspices of the Foundations of Computational Mathematics organisation.

Also in this series:

Simulating Hamiltonian Dynamics, *B. Leimkuhler and Sebastian Reich*

Collocation Methods for Volterra Integral and Related Functional Differential, *Hermann Brunner*

Topology for Computing, *Afra J. Zomorodian*

Scattered Data Approximation, *Holger Wendland*

Matrix Preconditioning Techniques and Applications, *Ke Chen*

Spectral Methods for Time-Dependent Problems, *Jan Hesthaven, Sigal Gottlieb and David Gottlieb*

The Mathematical Foundations of Mixing, *Rob Sturman, Julio M. Ottino and Stephen Wiggins*

Curve and Surface Reconstruction, *Tamal K. Dey*

Learning Theory, *Felipe Cucker and Ding Xuan Zhou*

Algebraic Geometry and Statistical Learning Theory

SUMIO WATANABE
Tokyo Institute of Technology

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521864671

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Watanabe, Sumio, 1959–
Algebraic geometry and statistical learning theory / Sumio Watanabe.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-86467-1 (hardback)
1. Computational learning theory – Statistical methods. 2. Geometry, Algebraic. I. Title.
Q325.7.W38 2009
006.3′1 – dc22 2009011366

ISBN 978-0-521-86467-1 hardback

Contents

Preface

In this book, we introduce a fundamental relation between algebraic geometry and statistical learning theory.

A lot of statistical models and learning machines used in information science, for example, mixtures of probability distributions, neural networks, hidden Markov models, Bayesian networks, stochastic context-free grammars, and topological data analysis, are not regular but singular, because they are non-identifiable and their Fisher information matrices are singular. In such models, knowledge to be discovered from examples corresponds to a singularity, hence it has been difficult to develop a mathematical method that enables us to understand statistical estimation and learning processes.

Recently, we established singular learning theory, in which four general formulas are proved for singular statistical models. Firstly, the log likelihood ratio function of any singular model can be represented by the common standard form even if it contains singularities. Secondly, the asymptotic behavior of the evidence or stochastic complexity is clarified, giving the result that the learning coefficient is equal to the maximum pole of the zeta function of a statistical model. Thirdly, there exist equations of states that express the universal relation of the Bayes quartet. We can predict Bayes and Gibbs generalization errors using Bayes and Gibbs training errors without any knowledge of the true distribution. And lastly, the symmetry of the generalization and training errors holds in the maximum likelihood and *a posteriori* estimators. If one-point estimation is applied to statistical learning, the generalization error is equal to the maximum value of a Gaussian process on a real analytic set.

This book consists of eight chapters. In Chapter 1, an outline of singular learning theory is summarized. The main formulas proved in this book are overviewed without mathematical preparation in advance. In Chapter 2, the definition of a singularity is introduced. Resolution of singularities is the essential theorem on which singular learning theory is constructed. In Chapter 3, several basic concepts in algebraic geometry are briefly explained: ring and ideal, correspondence between algebra and geometry, and projective spaces. The algorithm by which a resolution map is found using recursive blow-ups is also described. In Chapter 4, the relation between the singular integral and the zeta function of a singular statistical model is clarified, enabling some inequalities used in Chapter 6 to be proved. In Chapter 5, function-valued random variables are studied and convergence in law of empirical processes is proved. In Chapter 6, the four main formulas are proved: the standard form of the likelihood ratio function, the asymptotic expansion of the stochastic complexity, the equations of states in a Bayes quartet, and the symmetry of generalization and training errors in one-point estimation. In Chapters 7 and 8, applications of singular learning theory to information science are summarized and discussed.

This book involves several mathematical fields, for example, singularity theory, algebraic geometry, Schwartz distribution, and empirical processes. However, these mathematical concepts are introduced in each chapter for those who are unfamiliar with them. No specialized mathematical knowledge is necessary to read this book. The only thing the reader needs is a mathematical mind seeking to understand the real world.

The author would like to thank Professor Shun-ichi Amari for his encouragement of this research. Also the author would like to thank Professor Bernd Sturmfels for his many helpful comments on the study and this book.

In this book, the author tries to build a bridge between pure mathematics and real-world information science. It is expected that a new research field will be opened between algebraic geometry and statistical learning theory.

Sumio Watanabe