

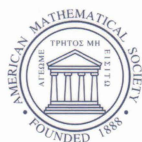
# CONTEMPORARY MATHEMATICS

443

## Prediction and Discovery

AMS-IMS-SIAM Joint Summer Research Conference  
Machine and Statistical Learning: Prediction and Discovery  
June 25–29, 2006  
Snowbird, Utah

Joseph Stephen Verducci  
Xiaotong Shen  
John Lafferty  
Editors



# CONTEMPORARY MATHEMATICS

---

443

## Prediction and Discovery

AMS-IMS-SIAM Joint Summer Research Conference  
Machine and Statistical Learning: Prediction and Discovery  
June 25–29, 2006  
Snowbird, Utah

Joseph Stephen Verducci  
Xiaotong Shen  
John Lafferty  
Editors



## Editorial Board

Dennis DeTurck, managing editor

George Andrews   Andreas Blass   Abel Klein

This volume contains the proceedings of an AMS-IMS-SIAM Joint Summer Research Conference on Machine and Statistical Learning, Prediction and Discovery, held in Snowbird, Utah, from June 25–29, 2006, with support from the National Science Foundation, grant DMS-0532564.

2000 *Mathematics Subject Classification*. Primary 62–04, 62–06.

---

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

---

### Library of Congress Cataloging-in-Publication Data

AMS-IMS-SIAM Joint Summer Research Conference, Machine and Statistical Learning, Prediction and Discovery (2006 : Snowbird, Utah)

Prediction and discovery : AMS-IMS-SIAM joint summer research conference on machine and statistical learning, prediction and discovery, June 25–29, 2006, Snowbird, Utah / Joseph Stephen Verducci et al., editors.

p. cm. — (Contemporary mathematics, ISSN 0271-4132 ; v. 443)

ISBN 978-0-8218-4195-2 (alk. paper)

Includes bibliographical references and index.

I. Machine theory—Congresses. I. Verducci, Joseph S. II. Shen, Xiaotong, 1964– III. Lafferty, John. IV. Title.

QA267.P7 2007

006.3'1—dc22

2007060788

---

**Copying and reprinting.** Material in this book may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to [reprint-permission@ams.org](mailto:reprint-permission@ams.org).

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

© 2007 by the American Mathematical Society. All rights reserved.

The American Mathematical Society retains all rights  
except those granted to the United States Government.

Copyright of individual articles may revert to the public domain 28 years  
after publication. Contact the AMS for copyright status of individual articles.

Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines  
established to ensure permanence and durability.

Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1    12 11 10 09 08 07

We would like to dedicate this volume  
to the memory of Leo Breiman,  
who brought us together.

## Contents

|   |     |
|---|-----|
| Introduction  | 1   |
| On transductive support vector machines<br>JUNHUI WANG, XIAOTONG SHEN, AND WEI PAN  | 7   |
| A note on robust kernel principal component analysis<br>XINWEI DENG, MING YUAN, AND AGUS SUDJIANTO                                    | 21  |
| The $L_q$ support vector machine<br>YUFENG LIU, HAO HELEN ZHANG, CHEOLWOO PARK, AND JEONGYOUN<br>AHN                                  | 35  |
| On multicategory truncated-hinge-loss support vector machines<br>YICHAO WU AND YUFENG LIU   | 49  |
| A robust hybrid of lasso and ridge regression<br>ART B. OWEN  | 59  |
| A gradient descent algorithm for LASSO<br>YONGDAI KIM, YUWON KIM, AND JINSEOG KIM   | 73  |
| Additive regression trees and smoothing splines - predictive modeling and<br>interpretation in data mining<br>BIN LI AND PREM K. GOEL | 83  |
| Estimation of atom prevalence for optimal prediction<br>ERNEST PARFAIT FOKOUÉ   | 103 |
| Precise statements of convergence for AdaBoost and arc-gv<br>CYNTHIA RUDIN, ROBERT E. SCHAPIRE, AND INGRID DAUBECHIES                 | 131 |
| Ensemble-learning by model-based spatial averaging<br>KEITH MARSOLO, SRINIVASAN PARTHASARATHY, MICHAEL TWA,<br>AND MARK BULLIMORE     | 147 |
| Automatic bias correction methods in semi-supervised learning<br>HUI ZOU, JI ZHU, SAHARON ROSSET, AND TREVOR HASTIE                   | 165 |
| Variable selection for model-based high-dimensional clustering<br>SIJIAN WANG AND JI ZHU  | 177 |
| Semi-supervised learning via constraints<br>WEI PAN AND XIAOTONG SHEN   | 193 |

|  |     |
|--|-----|
| Objective measures for association pattern analysis<br>MICHAEL STEINBACH, PANG-NING TAN, HUI XIONG, AND VIPIN<br>KUMAR | 205 |
|--|-----|

These proceedings feature some of the latest important results about machine learning based on methods originated in Computer Science and Statistics. In addition to papers discussing theoretical analysis of the performance of procedures for classification and prediction, the papers in this book cover novel versions of Support Vector Machines (SVM), Principal Component methods, Lasso prediction models, and Boosting and Clustering. Also included are applications such as multi-level spatial models for diagnosis of eye disease, hyperclique methods for identifying protein interactions, robust SVM models for detection of fraudulent banking transactions, etc.

This book should be of interest to researchers who want to learn about the various new directions that the field is taking, to graduate students who want to find a useful and exciting topic for their research or learn the latest techniques for conducting comparative studies, and to engineers and scientists who want to see examples of how to modify the basic high-dimensional methods to apply to real world applications with special conditions and constraints.

ISBN 978-0-8218-4195-2



9 780821 841952

**CONM/443**

*AMS on the Web*  
**www.ams.org**