Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Support Vector Machines

Optimization Based Theory, Algorithms, and Extensions

Naiyang Deng
Yingjie Tian
Chunhua Zhang

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Support Vector Machines

Optimization Based Theory, Algorithms, and Extensions

Naiyang Deng
Yingjie Tian
Chunhua Zhang

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
Version Date: 20121112

International Standard Book Number: 978-1-4398-5792-2 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Support Vector Machines

Optimization Based Theory, Algorithms, and Extensions

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

SERIES EDITOR

Vipin Kumar
University of Minnesota
Department of Computer Science and Engineering
Minneapolis, Minnesota, U.S.A.

AIMS AND SCOPE

This series aims to capture new developments and applications in data mining and knowledge discovery, while summarizing the computational tools and techniques useful in data analysis. This series encourages the integration of mathematical, statistical, and computational methods and techniques through the publication of a broad range of textbooks, reference works, and handbooks. The inclusion of concrete examples and applications is highly encouraged. The scope of the series includes, but is not limited to, titles in the areas of data mining and knowledge discovery methods and applications, modeling, algorithms, theory and foundations, data and knowledge visualization, data mining systems and tools, and privacy and security issues.

PUBLISHED TITLES

ADVANCES IN MACHINE LEARNING AND DATA MINING FOR ASTRONOMY
Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, and Ashok N. Srivastava

BIOLOGICAL DATA MINING
Jake Y. Chen and Stefano Lonardi

COMPUTATIONAL METHODS OF FEATURE SELECTION
Huan Liu and Hiroshi Motoda

CONSTRAINED CLUSTERING: ADVANCES IN ALGORITHMS, THEORY, AND APPLICATIONS
Sugato Basu, Ian Davidson, and Kiri L. Wagstaff

CONTRAST DATA MINING: CONCEPTS, ALGORITHMS, AND APPLICATIONS
Guozhu Dong and James Bailey

DATA CLUSTERING IN C++: AN OBJECT-ORIENTED APPROACH
Guojun Gan

DATA MINING FOR DESIGN AND MARKETING
Yukio Ohsawa and Katsutoshi Yada

DATA MINING WITH R: LEARNING WITH CASE STUDIES
Luís Torgo

FOUNDATIONS OF PREDICTIVE ANALYTICS
James Wu and Stephen Coggeshall

GEOGRAPHIC DATA MINING AND KNOWLEDGE DISCOVERY, SECOND EDITION
Harvey J. Miller and Jiawei Han

HANDBOOK OF EDUCATIONAL DATA MINING
Cristóbal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan S.J.d. Baker

INFORMATION DISCOVERY ON ELECTRONIC HEALTH RECORDS
Vagelis Hristidis

INTELLIGENT TECHNOLOGIES FOR WEB APPLICATIONS
Priti Srinivas Sajja and Rajendra Akerkar

INTRODUCTION TO PRIVACY-PRESERVING DATA PUBLISHING: CONCEPTS AND TECHNIQUES
Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu

KNOWLEDGE DISCOVERY FOR COUNTERTERRORISM AND LAW ENFORCEMENT
David Skillicorn

KNOWLEDGE DISCOVERY FROM DATA STREAMS
João Gama

MACHINE LEARNING AND KNOWLEDGE DISCOVERY FOR ENGINEERING SYSTEMS HEALTH MANAGEMENT
Ashok N. Srivastava and Jiawei Han

MINING SOFTWARE SPECIFICATIONS: METHODOLOGIES AND APPLICATIONS
David Lo, Siau-Cheng Khoo, Jiawei Han, and Chao Liu

MULTIMEDIA DATA MINING: A SYSTEMATIC INTRODUCTION TO CONCEPTS AND THEORY
Zhongfei Zhang and Ruofei Zhang

MUSIC DATA MINING
Tao Li, Mitsunori Ogihara, and George Tzanetakis

NEXT GENERATION OF DATA MINING
Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, and Vipin Kumar

RELATIONAL DATA CLUSTERING: MODELS, ALGORITHMS, AND APPLICATIONS
Bo Long, Zhongfei Zhang, and Philip S. Yu

SERVICE-ORIENTED DISTRIBUTED KNOWLEDGE DISCOVERY
Domenico Talia and Paolo Trunfio

SPECTRAL FEATURE SELECTION FOR DATA MINING
Zheng Alan Zhao and Huan Liu

STATISTICAL DATA MINING USING SAS APPLICATIONS, SECOND EDITION
George Fernandez

SUPPORT VECTOR MACHINES: OPTIMIZATION BASED THEORY, ALGORITHMS, AND EXTENSIONS
Naiyang Deng, Yingjie Tian, and Chunhua Zhang

TEMPORAL DATA MINING
Theophano Mitsa

TEXT MINING: CLASSIFICATION, CLUSTERING, AND APPLICATIONS
Ashok N. Srivastava and Mehran Sahami

THE TOP TEN ALGORITHMS IN DATA MINING
Xindong Wu and Vipin Kumar

UNDERSTANDING COMPLEX DATASETS: DATA MINING WITH MATRIX DECOMPOSITIONS
David Skillicorn

Dedicated to my beloved wife Meifang

Naiyang Deng

Dedicated to my dearest father Mingran Tian

Yingjie Tian

Dedicated to my husband Xingang Xu and my son Kaiwen Xu

Chunhua Zhang

Preface

Support vector machines (SVMs), which were introduced by Vapnik in the early 1990s, have proven effective and promising techniques for data mining. SVMs have recently made breakthroughs and advances in their theoretical studies and implementations of algorithms. They have been successfully applied in many fields such as text categorization, speech recognition, remote sensing image analysis, time series forecasting, information security, and so forth.

SVMs, having their roots in Statistical Learning Theory (SLT) and optimization methods, have become powerful tools to solve the problems of machine learning with finite training points and to overcome some traditional difficulties such as the "curse of dimensionality", "over-fitting", and so forth. Their theoretical foundation and implementation techniques have been established and SVMs are gaining quick popularity due to their many attractive features: nice mathematical representations, geometrical explanations, good generalization abilities, and promising empirical performance. Some SVM monographs, including more sophisticated ones such as Cristianini & Shawe-Taylor [39] and Scholkopf & Smola [124], have been published.

We have published two books in Chinese about SVMs in Science Press of China since 2004 [42, 43], which attracted widespread interest and received favorable comments in China. After several years of research and teaching, we decided to rewrite the books and add new research achievements. The starting point and focus of the book is optimization theory, which is different from other books on SVMs in this respect. Optimization is one of the pillars on which SVMs are built, so it makes a lot of sense to consider them from this point of view.

This book introduces SVMs systematically and comprehensively. We place emphasis on the readability and the importance of perception on a sound understanding of SVMs. Prior to systematical and rigorous discourses, concepts are introduced graphically, and the methods and conclusions are proposed by direct inspection or with visual explanation. Particularly, for some important concepts and algorithms we try our best to give clearly geometric interpretations that are not depicted in the literature, such as Crammer-Singer SVM for multiclass classification problems.

We give details on classification problems and regression problems that are the two main components of SVMs. We formated this book uniformly by using the classification problem as the principal axis and converting the regression problem to the classification problem. The book is organized as follows. In Chapter 1 the optimization fundamentals are introduced. The convex programming encompassing traditional convex optimization (Sections 1.1–1.3) and conic programming (Sections 1.4-1.5). Sections 1.1–1.3 are necessary background for the later chapters. For beginners, Sections 1.4 and 1.5 (marked with an asterisk *) can be skipped since they are used only in Subsections 8.4.3 and 8.8.3 of Chapter 8, and are mainly served for further research. Support vector machines begin from Chapter 2 starting from linear classification problems. Based on the maximal margin principal, the basic linear support vector classification is derived visually in Chapter 2. Linear support vector regression is established in Chapter 3. The kernel theory, which is the key of extension of basic SVMs and the foundation for solving nonlinear problems, together with the general classification and regression problems, are discussed in Chapter 4. Starting with statistical interpretation of the maximal margin method, statistical learning theory, which is the groundwork of SVMs, is studied in Chapter 5. The model construction problems, which are very useful in practical applications, are discussed in Chapter 6. The implementations of several prevailing SVM's algorithms are introduced in Chapter 7. Finally, the variations and extensions of SVMs including multiclass classification, semisupervised classification, knowledge-based classification, Universum classification, privileged classification, robust classification, multi-instance classification, and multi-label classification are covered in Chapter 8.

The contents of this book comprise our research achievements. A precise and concise interpretation of statistical leaning theory for C-support vector classification (C-SVC) is given in Chapter 5 which imbues the parameter C with a new meaning. From our achievements the following results of SVMs are also given: the regularized twin SVMs for binary classification problems, the SVMs for solving multi-classification problems based on the idea of ordinal regression, the SVMs for semisupervised problems by means of constructing second order cone programming or semidefinite programming models, and the SVMs for problems with perturbations.

Potential readers include those who are beginners in the SVM and those who are interested in solving real-world problems by employing SVMs, and those who will conduct more comprehensive study of SVMs.

We are indebted to all the people who have helped in various ways. We would like to say special thanks to Dr. Hang Li, Chief Scientist of Noah's Ark Lab of Huawei Technologies, academicians Zhiming Ma and Yaxiang Yuan of Chinese Academy of Sciences, Dr. Mingren Shi of University of Western Australia, Prof. Changyu Wang and Prof. Yiju Wang of Qufu Normal University, Prof. Zunquan Xia and Liwei Zhang of Dalian University of Technology, Prof. Naihua Xiu of Beijing Jiaotong University, Prof. Yanqin Bai of Shanghai University, and Prof. Ling Jing of China Agricultural University for their valuable suggestions. Our gratitude goes also to Prof. Xiangsun Zhang and Prof. Yong Shi of Chinese Academy of Sciences, and Prof. Shuzhong Zhang of The Chinese University of Hong Kong for their great help and support. We appreciate assistance from the members of our workshop — Dr. Zhixia Yang, Dr. Kun Zhao, Dr. Yongcui Wang, Dr. Xiaojian Shao, Dr. Ruxin Qin, Dr. Yuanhai Shao, Dr. Junyan Tan, Ms. Yanmei Zhao, Ms. Tingting Gao, and Ms. Yuxin Li.

Finally, we would like acknowledge a number of funding agencies that provided their generous support to our research activities on this book. They are the Publishing Foundation of The Ministry of Science and Technology of China, and the National Natural Science Foundation of China, including the innovative group grant "Data Mining and Intelligent Knowledge Management" (♯70621001, ♯70921061); the general project " Knowledge Driven Support Vector Machines Theory, Algorithms and Applications" (♯ 11271361); the general project "Models and Algorithms for Support Vector Machines with Adaptive Norms" (♯ 11201480); the general project "The Optimization Methods in Multi-label Multi-instance Learning and its Applications" (♯10971223); the general project "The Optimization Methods of Kernel Matrix Learning and its Applications in Bioinformatics" (♯11071252); the CAS/SAFEA International Partnership Program for Creative Research Teams; the President Fund of GUCAS; and the National Technology Support Program 2009BAH42B02.

List of Symbols

R	real numbers
$x \in R^n$	input and Euclidian space
$y \in \mathcal{Y}$	output and set of output
(x_i, y_i)	the ith training point
$T = \{(x_1, y_1), \cdots, (x_l, y_l)\}$	training set
l	number of training points
$[x]_i, [x_i]_j$	the ith component of the vector x, the jth component of the vector $[x_i]_j$
$\mathrm{x} = \Phi(x)$	vector in Hilbert space and mapping from input space into Hilbert space
$[\mathrm{x}]_i, [\mathrm{x}_i]_j$	the ith component of vector x, the jth component of vector x_i
$(x \cdot x')$, $(\mathrm{x} \cdot \mathrm{x}')$	inner product between x and x', inner product between x and x′
$\mathcal{H}$	Hilbert space
w	weight vector in R^n
w_i	the ith component of w
w	weight vector in $\mathcal{H}$
w_i	the ith component of w
b	threshold
$K(x, x')$	kernel function $(\Phi(x) \cdot \Phi(x'))$
K	kernel matrix (Gram matrix)
$\lVert \cdot \rVert_p$	p-norm
$\lVert \cdot \rVert$	2-norm
$\lVert \cdot \rVert_1$	1-norm
h	VC dimension
C	penalty parameter
ξ	vector of slack variables
ξ_i	the ith component of ξ
α	dual variables, vector of Lagrange multipliers
α_i	the ith component of α
β	dual variables, vector of Lagrange multipliers
β_i	the ith component of β
$P(\cdot)$	possibility distribution or possibility

Contents

List of Figures xvii

List of Tables xxi

Preface xxiii

List of Symbols xxvii

1 Optimization 1

1.1 Optimization Problems in Euclidian Space 1

1.1.1 An example of optimization problems 1

1.1.2 Optimization problems and their solutions 3

1.1.3 Geometric interpretation of optimization problems . . 4

1.2 Convex Programming in Euclidean Space 5

1.2.1 Convex sets and convex functions 6

1.2.1.1 Convex sets 6

1.2.1.2 Convex functions 7

1.2.2 Convex programming and their properties 8

1.2.2.1 Convex programming problems 8

1.2.2.2 Basic properties 9

1.2.3 Duality theory 12

1.2.3.1 Derivation of the dual problem 12

1.2.3.2 Duality theory 13

1.2.4 Optimality conditions 15

1.2.5 Linear programming 16

1.3 Convex Programming in Hilbert Space 18

1.3.1 Convex sets and Fréchet derivative 18

1.3.2 Convex programming problems 19

1.3.3 Duality theory 20

1.3.4 Optimality conditions 20

*1.4 Convex Programming with Generalized Inequality Constraints in Euclidian Space 21

1.4.1 Convex programming with generalized inequality constraints 21

1.4.1.1 Cones 21

1.4.1.2 Generalized inequalities 21

1.4.1.3 Convex programming with generalized inequality constraints . . . 22
1.4.2 Duality theory . . . 23
1.4.2.1 Dual cones . . . 23
1.4.2.2 Derivation of the dual problem . . . 23
1.4.2.3 Duality theory . . . 25
1.4.3 Optimality conditions . . . 26
1.4.4 Second-order cone programming . . . 27
1.4.4.1 Second-order cone programming and its dual problem . . . 27
1.4.4.2 Software for second-order cone programming 30
1.4.5 Semidefinite programming . . . 31
1.4.5.1 Semidefinite programming and its dual problem . . . 31
1.4.5.2 Software for semidefinite programming . . . 35
*1.5 Convex Programming with Generalized Inequality Constraints in Hilbert Space . . . 36
1.5.1 K-convex function and Fréchet derivative . . . 36
1.5.2 Convex programming . . . 36
1.5.3 Duality theory . . . 37
1.5.4 Optimality conditions . . . 38

2 Linear Classification **41**
2.1 Presentation of Classification Problems . . . 41
2.1.1 A sample (diagnosis of heart disease) . . . 41
2.1.2 Classification problems and classification machines . . 43
2.2 Support Vector Classification (SVC) for Linearly Separable Problems . . . 45
2.2.1 Maximal margin method . . . 45
2.2.1.1 Derivation of the maximal margin method . 45
2.2.1.2 Properties of the maximal margin method . 48
2.2.2 Linearly separable support vector classification 50
2.2.2.1 Relationship between the primal and dual problems . . . 50
2.2.2.2 Linearly separable support vector classification . . . 53
2.2.3 Support vector . . . 54
2.3 Linear C-Support Vector Classification . . . 56
2.3.1 Maximal margin method . . . 56
2.3.1.1 Derivation of the maximal margin method . 56
2.3.1.2 Properties of the maximal margin method . 57
2.3.2 Linear C-support vector classification . . . 59
2.3.2.1 Relationship between the primal and dual problems . . . 59
2.3.2.2 Linear C-support vector classification 60