Preface

Every mathematical discipline goes
through three periods of development:
the naive, the formal, and the critical.
David Hilbert

The goal of this book is to explain the principles that made support vector machines (SVMs) a successful modeling and prediction tool for a variety of applications. We try to achieve this by presenting the basic ideas of SVMs together with the latest developments and current research questions in a unified style. In a nutshell, we identify at least three reasons for the success of SVMs: their ability to learn well with only a very small number of free parameters, their robustness against several types of model violations and outliers, and last but not least their computational efficiency compared with several other methods.

Although there are several roots and precursors of SVMs, these methods gained particular momentum during the last 15 years since Vapnik (1995, 1998) published his well-known textbooks on statistical learning theory with a special emphasis on support vector machines. Since then, the field of machine learning has witnessed intense activity in the study of SVMs, which has spread more and more to other disciplines such as statistics and mathematics. Thus it seems fair to say that several communities are currently working on support vector machines and on related kernel-based methods. Although there are many interactions between these communities, we think that there is still room for additional fruitful interaction and would be glad if this textbook were found helpful in stimulating further research. Many of the results presented in this book have previously been scattered in the journal literature or are still under review. As a consequence, these results have been accessible only to a relatively small number of specialists, sometimes probably only to people from one community but not the others. In view of the importance of SVMs for statistical machine learning, we hope that the unified presentation given here will make these results more accessible to researchers and users from different

communities (e.g.; from the fields of statistics, mathematics, computer science, bioinformatics, data and text mining, and engineering).

As in most monographs, the selection of topics treated in this textbook is biased for several reasons. We have of course focused on those that particularly interest us and those that we have been working on during the last decade. We also decided to concentrate on some important and selected topics, so for these topics we can offer not only the results but also the proofs. This is in contrast to some other textbooks on SVMs or statistical machine learning in general, but we try to follow the path described by Devroye *et al.* (1996) and Györfi *et al.* (2002). Moreover, some topics, such as the robustness properties of SVMs, a detailed treatment of loss functions and reproducing kernel Hilbert spaces, recent advances in the statistical analysis of SVMs, and the relationship between good learning properties and good robustness properties such as a bounded influence function and a bounded maxbias, are not covered by other currently available books on SVMs. On the other hand, the area of statistical machine learning is nowadays so rich and progressing so rapidly that covering all aspects in detail in a single book hardly seems possible. The consequence is of course that several important and interesting topics of SVMs and related methods are not covered in this monograph. This includes, for example, SVMs for anomaly detection, kernel principal component analysis, kernel-based independence measures, structured estimation, recent progress in computational algorithms, boosting, Bayesian approaches, and the analysis of time series or text data. A reader interested in these topics will get useful information in the books by Vapnik (1995, 1998), Cristianini and Shawe-Taylor (2000), Hastie *et al.* (2001), Schölkopf and Smola (2002), Shawe-Taylor and Cristianini (2004), and Bishop (2006), among others. Moreover, many of the most recent developments can be found in journals such as *Journal of Machine Learning Research*, *Machine Learning*, and *Annals of Statistics*, or in the proceedings to conferences such as *NIPS* or *COLT*.

The process of writing this book took about four years. Springer asked one of us (A.C.), after a talk on robustness properties of SVMs, whether he was willing to write a book on this topic. After a few weeks to convince Ingo to write a joint textbook, our plan was to write a very condensed book of about 200–250 pages within one-and-a-half years. However, this soon turned out to be unrealistic because we aimed to include some of our own research results that were partially written simultaneously to the book. Moreover, we totally underestimated the richness of the available literature on SVMs and the field's speed of progress.

Acknowledgments

First of all, we would like to thank our wives, Anne and Wiebke, for their permanent support and understanding during the entire process of writing

the book. Without their patience it would have been completely impossible for us to write this book.

My (I.S.) special thanks go to Dieter Eilts, who played so sensationally at the European Championship in 1996 that the semi-final was the classic England vs. Germany. This in turn caused Bob Williamson, who stayed during the time in London, to flee from the crowded city into a secondhand bookstore, where he found a book on entropy numbers by my Ph.D. adviser Prof. Bernd Carl. Bob, who was working closely with Alex Smola and Bernhard Schölkopf, contacted Bernd Carl, and after some time Alex and Bernhard were visiting us in Jena. Inspired by their visit, I decided to work on SVMs after finishing my Ph.D. thesis, and it seems fair to say that without Dieter Eilts (and, of course, the other persons mentioned) this book would have never been written. I also thank Prof. Andreas Defant and Prof. Klaus Floret for teaching me to look behind the formulas of mathematics, and Prof. Bernd Carl and Prof. Erich Novak, who greatly supported me when I changed my focus of research to SVMs.

My (A.C.) special thanks go to Prof. Dr. Ursula Gather from the University of Dortmund for bringing robust and non-parametric statistics, rates of convergence, and other topics of statistics to my attention during my studies and later on when I was a member of the Collaborative Research Center 475 on "Reduction of Complexity in Multivariate Data Structures." Her experience and her support were always very valuable to me.

We would like to thank the following friends and colleagues who have, over the years, collaborated with us on various aspects of the work described here, provided numerous helpful discussions, and brought several references to our attention: F. Alexander, M. Anghel, P. Bartlett, J. Beirlant, O. Bousquet, C. Croux, I. Daubechies, P. L. Davies, M. Debruyne, H. Dette, U. Einmahl, L. Fernholz, L. Rosasco, P. Filzmoser, U. Gather, M. Hallin, X. He, M. Hein, A. Hinrichs, J. Howse, M. Hubert, D. Hush, T. Joachims, M. Kohl, V. Koltchinskii, N. List, G. Lugosi, S. Mendelson, S. Morgenthaler, K. Mosler, S. Mukherjee, H. Oja, D. Paindaveine, S. Portnoy, H. Rieder, E. Ronchetti, P. J. Rousseeuw, P. Ruckdeschel, S. Rüping, M. Salibian-Barrera, B. Schölkopf, C. Scovel, H. U. Simon, S. Smale, A. Smola, J. Suykens, V. Temlyakov, J. Teugels, J. Theiler, A. Tsybakov, D. Tyler, S. Van Aelst, A. Van Messem, U. von Luxburg, C. Weihs, B. Williamson, T. Zhang, and D. X. Zhou.

We deeply thank F. Alexander, M. Anghel, U. Einmahl, U. Gather, M. Hein, A. Hinrichs, M. Kohl, P. Ruckdeschel, C. Scovel, J. Theiler, and A. Van Messem, who critically read parts of the book and made many valuable suggestions.

We thank Springer and, in particular, Executive Editor John Kimmel and Frank Glanz for their support and constant advice and the editors of the Springer Series on Information Science and Statistics, especially Prof. Bernhard Schölkopf, for their support. We would also like to thank many unknown

reviewers who read one or more chapters of earlier versions of the manuscript for constructive criticism that led to considerable improvements.

Finally, we thank the Los Alamos LDRD program for basic research for grants and support for the first author. We thank the Deutsche Forschungsgesellschaft (DFG); the Collaborative Research Center SFB 475 and the interdisciplinary research program DoMuS, both at the University of Dortmund in Germany; and the Fonds voor Wetenschappelijk Onderzoek (FWO) in Belgium for grants and additional support for the second author.

Ingo Steinwart
Los Alamos National Laboratory
Group CCS-3
Los Alamos, NM, USA

Andreas Christmann
University of Bayreuth
Department of Mathematics
Bayreuth, GERMANY

Spring 2008

http://www.springer.com/978-0-387-77241-7

Support Vector Machines
Steinwart, I.; Christmann, A.
2008, XVI, 601 p., Hardcover
ISBN: 978-0-387-77241-7