Preface

One of the most fortunate situations a scientist can encounter is to enter a field in its infancy. There is a large choice of topics to work on, and many of the issues are conceptual rather than merely technical. Over the last seven years, we have had the privilege to be in this position with regard to the field of Support Vector Machines (SVMs). We began working on our respective doctoral dissertations in 1994 and 1996. Upon completion, we decided to combine our efforts and write a book about SVMs. Since then, the field has developed impressively, and has to an extent been transformed. We set up a website that quickly became the central repository for the new community, and a number of workshops were organized by various researchers. The scope of the field has now widened significantly, both in terms of new algorithms, such as kernel methods different to SVMs, and in terms of a deeper theoretical understanding being gained. It has become clear that kernel methods provide a framework for tackling some rather profound issues in machine learning theory. At the same time, successful applications have demonstrated that SVMs not only have a more solid foundation than artificial neural networks, but are able to serve as a replacement for neural networks that perform as well or better, in a wide variety of fields. Standard neural network and pattern recognition textbooks have now started including chapters on SVMs and kernel PCA (for instance, [235, 153]).

While these developments took place, we were trying to strike a balance between pursuing exciting new research, and making progress with the slowly growing manuscript of this book. In the two and a half years that we worked on the book, we faced a number of lessons that we suspect everyone writing a scientific monograph — or any other book — will encounter. First, writing a book is more work than you think, even with two authors sharing the work in equal parts. Second, our book got longer than planned. Once we exceeded the initially planned length of 500 pages, we got worried. In fact, the manuscript kept growing even after we stopped writing new chapters, and began polishing things and incorporating corrections suggested by colleagues. This was mainly due to the fact that the book deals with a fascinating new area, and researchers keep adding fresh material to the body of knowledge. We learned that there is no asymptotic regime in writing such a book — if one does not stop, it will grow beyond any bound — unless one starts cutting. We therefore had to take painful decisions to leave out material that we originally thought should be in the book. Sadly, and this is the third point, the book thus contains less material than originally planned, especially on the subject of theoretical developments. We sincerely apologize to all researchers who feel that their contributions should have been included — the book is certainly biased towards our own work, and does not provide a fully comprehensive overview of the field. We did, however, aim to provide all the necessary concepts and ideas to enable a reader equipped with some basic mathematical knowledge to enter the engaging world of machine learning, using theoretically well-founded kernel algorithms, and to understand and apply the powerful algorithms that have been developed over the last few years.

The book is divided into three logical parts. Each part consists of a brief introduction and a number of technical chapters. In addition, we include two appendices containing addenda, technical details, and mathematical prerequisites. Each chapter begins with a short discussion outlining the contents and prerequisites; for some of the longer chapters, we include a graph that sketches the logical structure and dependencies between the sections. At the end of most chapters, we include a set of problems, ranging from simple exercises (marked by •) to hard ones (•••); in addition, we describe open problems and questions for future research (◦◦◦).[1] The latter often represent worthwhile projects for a research publication, or even a thesis. References are also included in some of the problems. These references contain the solutions to the associated problems, or at least significant parts thereof.

The overall structure of the book is perhaps somewhat unusual. Rather than presenting a logical progression of chapters building upon each other, we occasionally touch on a subject briefly, only to revisit it later in more detail. For readers who are used to reading scientific monographs and textbooks from cover to cover, this will amount to some redundancy. We hope, however, that some readers, who are more selective in their reading habits (or less generous with their time), and only look at those chapters that they are interested in, will benefit. Indeed, nobody is expected to read every chapter. Some chapters are fairly technical, and cover material included for reasons of completeness. Other chapters, which are more relevant to the central subjects of the book, are kept simpler, and should be accessible to undergraduate students.

In a way, this book thus contains several books in one. For instance, the first chapter can be read as a standalone "executive summary" of Support Vector and kernel methods. This chapter should also provide a fast entry point for practitioners. Someone interested in applying SVMs to a pattern recognition problem might want to read Chapters 1 and 7 only. A reader thinking of building their own SVM implementation could additionally read Chapter 10, and parts of Chapter 6. Those who would like to get actively involved in research aspects of kernel methods, for example by "kernelizing" a new algorithm, should probably read at least Chapters 1 and 2. A one-semester undergraduate course on learning with kernels could include the material of Chapters 1, 2.1–2.3, 3.1–3.2, 5.1–5.2, 6.1–6.3, 7. If there is more

1. We suggest that authors post their solutions on the book website www.learning-with-kernels.org.

time, one of the Chapters 14, 16, or 17 can be added, or 4.1–4.2. A graduate course could additionally deal with the more advanced parts of Chapters 3, 4, and 5. The remaining chapters provide ample material for specialized courses and seminars.

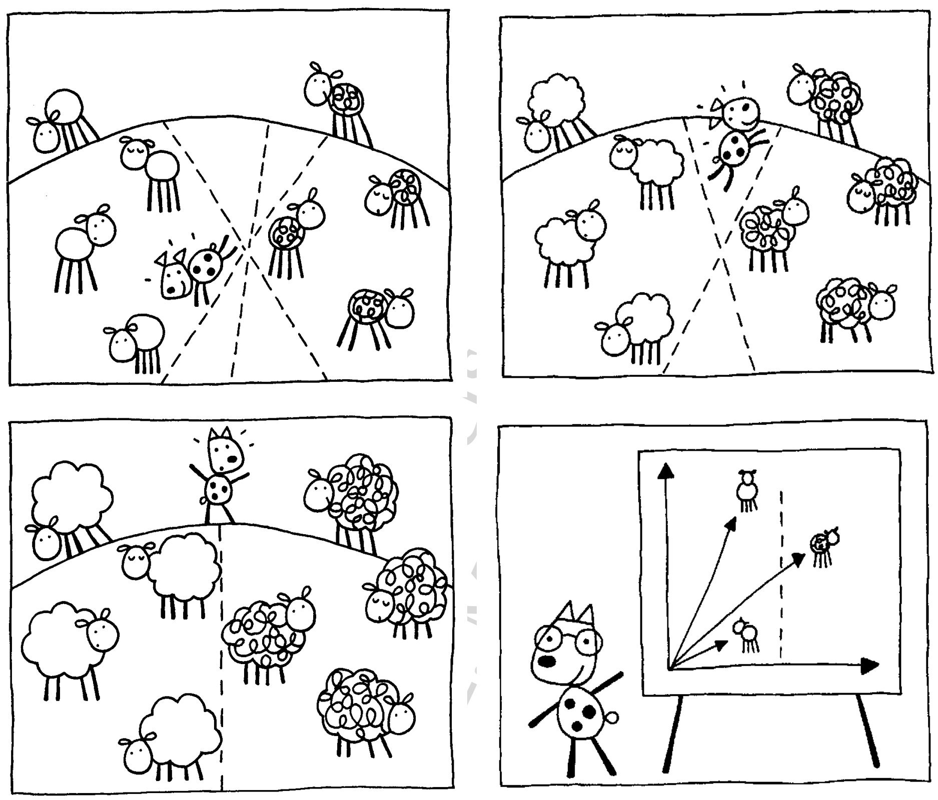
As a general time-saving rule, we recommend reading the first chapter and then jumping directly to the chapter of particular interest to the reader. Chances are that this will lead to a chapter that contains references to the earlier ones, which can then be followed as desired. We hope that this way, readers will inadvertently be tempted to venture into some of the less frequented chapters and research areas. Explore this book; there is a lot to find, and much more is yet to be discovered in the field of learning with kernels.

We conclude the preface by thanking those who assisted us in the preparation of the book. Our first thanks go to our first readers. Chris Burges, Arthur Gretton, and Bob Williamson have read through various versions of the book, and made numerous suggestions that corrected or improved the material. A number of other researchers have proofread various chapters. We would like to thank Matt Beal, Daniel Berger, Olivier Bousquet, Ben Bradshaw, Nicolò Cesa-Bianchi, Olivier Chapelle, Dennis DeCoste, Andre Elisseeff, Anita Faul, Arnulf Graf, Isabelle Guyon, Ralf Herbrich, Simon Hill, Dominik Janzing, Michael Jordan, Sathiya Keerthi, Neil Lawrence, Ben O'Loghlin, Ulrike von Luxburg, Davide Mattera, Sebastian Mika, Natasa Milic-Frayling, Marta Milo, Klaus Müller, Dave Musicant, Fernando Pérez Cruz, Ingo Steinwart, Mike Tipping, and Chris Williams.

In addition, a large number of people have contributed to this book in one way or another, be it by sharing their insights with us in discussions, or by collaborating with us on some of the topics covered in the book. In many places, this strongly influenced the presentation of the material. We would like to thank Dimitris Achlioptas, Luís Almeida, Shun-Ichi Amari, Peter Bartlett, Jonathan Baxter, Tony Bell, Shai Ben-David, Kristin Bennett, Matthias Bethge, Chris Bishop, Andrew Blake, Volker Blanz, Léon Bottou, Paul Bradley, Chris Burges, Heinrich Bülthoff, Olivier Chapelle, Nello Cristianini, Corinna Cortes, Cameron Dawson,Tom Dietterich, André Elisseeff, Oscar de Feo, Federico Girosi, Thore Graepel, Isabelle Guyon, Patrick Haffner, Stefan Harmeling, Paul Hayton, Markus Hegland, Ralf Herbrich, Tommi Jaakkola, Michael Jordan, Jyrki Kivinen, Yann LeCun, Chi-Jen Lin, Gabor Lugosi, Olvi Mangasarian, Laurent Massoulie, Sebastian Mika, Sayan Mukherjee, Klaus Müller, Noboru Murata, Nuria Oliver, John Platt, Tomaso Poggio, Gunnar Rätsch, Sami Romdhani, Rainer von Sachs, Christoph Schnörr, Matthias Seeger, John Shawe-Taylor, Kristy Sim, Patrice Simard, Stephen Smale, Sara Solla, Lionel Tarassenko, Lily Tian, Mike Tipping, Alexander Tsybakov, Lou van den Dries, Santosh Venkatesh, Thomas Vetter, Chris Watkins, Jason Weston, Chris Williams, Bob Williamson, Andreas Ziehe, Alex Zien, and Tong Zhang.

Next, we would like to extend our thanks to the research institutes that allowed us to pursue our research interests and to dedicate the time necessary for writing the present book; these are AT&T / Bell Laboratories (Holmdel), the Australian National University (Canberra), Biowulf Technologies (New York), GMD FIRST (Berlin), the Max-Planck-Institute for Biological Cybernetics (Tübingen), and Microsoft Research (Cambridge). We are grateful to Doug Sery from MIT Press for continuing support and encouragement during the writing of this book. We are, moreover, indebted to funding from various sources; specifically, from the Studienstiftung des deutschen Volkes, the Deutsche Forschungsgemeinschaft, the Australian Research Council, and the European Union.

Finally, special thanks go to Vladimir Vapnik, who introduced us to the fascinating world of statistical learning theory.



. . . the story of the sheep dog who was herding his sheep, and serendipitously invented both large margin classification and Sheep Vectors. . .

Illustration by Ana Martín Larrañaga