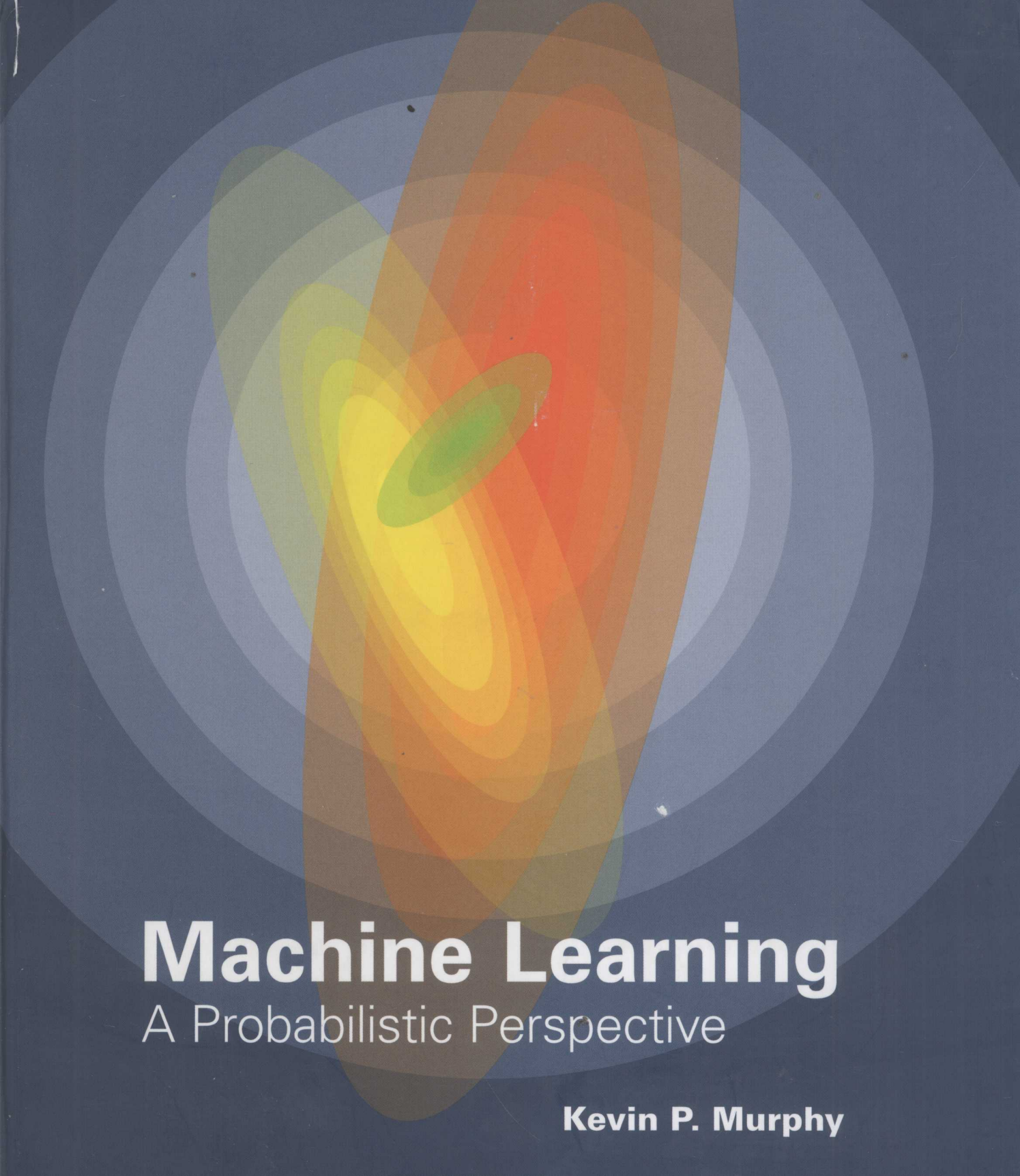
Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

Machine Learning
A Probabilistic Perspective

Kevin P. Murphy

The MIT Press
Cambridge, Massachusetts
London, England

For information about special quantity discounts, please email special_sales@mitpress.mit.edu

This book was set in the LaTeX programming language by the author. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Information

Murphy, Kevin P.
Machine learning : a probabilistic perspective / Kevin P. Murphy.
p. cm. — (Adaptive computation and machine learning series)
Includes bibliographical references and index.
ISBN 978-0-262-01802-9 (hardcover : alk. paper)
1. Machine learning. 2. Probabilities. I. Title.
Q325.5.M87 2012
006.3'1—dc23
2012004558

10 9 8 7 6 5 4 3

Machine Learning: A Probabilistic Perspective

Adaptive Computation and Machine Learning
Thomas Dietterich, Editor
Christopher Bishop, David Heckerman, Michael Jordan, and Michael Kearns, Associate Editors

Bioinformatics: The Machine Learning Approach, Pierre Baldi and Søren Brunak
Reinforcement Learning: An Introduction, Richard S. Sutton and Andrew G. Barto
Graphical Models for Machine Learning and Digital Communication, Brendan J. Frey
Learning in Graphical Models, Michael I. Jordan
Causation, Prediction, and Search, second edition, Peter Spirtes, Clark Glymour, and Richard Scheines
Principles of Data Mining, David Hand, Heikki Mannila, and Padhraic Smyth
Bioinformatics: The Machine Learning Approach, second edition, Pierre Baldi and Søren Brunak
Learning Kernel Classifiers: Theory and Algorithms, Ralf Herbrich
Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, Bernhard Schölkopf and Alexander J. Smola
Introduction to Machine Learning, Ethem Alpaydin
Gaussian Processes for Machine Learning, Carl Edward Rasmussen and Christopher K.I. Williams
Semi-Supervised Learning, Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, Eds.
The Minimum Description Length Principle, Peter D. GrÃijnwald
Introduction to Statistical Relational Learning, Lise Getoor and Ben Taskar, Eds.
Probabilistic Graphical Models: Principles and Techniques, Daphne Koller and Nir Friedman
Introduction to Machine Learning, second edition, Ethem Alpaydin
Boosting: Foundations and Algorithms, Robert E. Schapire and Yoav Freund
Machine Learning: A Probabilistic Perspective, Kevin P. Murphy
Foundations of Machine Learning, Mehryar Mohri, Afshin Rostami, and Ameet Talwalker

This book is dedicated to Alessandro, Michael and Stefano,
and to the memory of Gerard Joseph Murphy.

Preface

Introduction

With the ever increasing amounts of data in electronic form, the need for automated methods for data analysis continues to grow. The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest. Machine learning is thus closely related to the fields of statistics and data mining, but differs slightly in terms of its emphasis and terminology. This book provides a detailed introduction to the field, and includes worked examples drawn from application domains such as molecular biology, text processing, computer vision, and robotics.

Target audience

This book is suitable for upper-level undergraduate students and beginning graduate students in computer science, statistics, electrical engineering, econometrics, or any one else who has the appropriate mathematical background. Specifically, the reader is assumed to already be familiar with basic multivariate calculus, probability, linear algebra, and computer programming. Prior exposure to statistics is helpful but not necessary.

A probabilistic approach

This books adopts the view that the best way to make machines that can learn from data is to use the tools of probability theory, which has been the mainstay of statistics and engineering for centuries. Probability theory can be applied to any problem involving uncertainty. In machine learning, uncertainty comes in many forms: what is the best prediction (or decision) given some data? what is the best model given some data? what measurement should I perform next? etc.

The systematic application of probabilistic reasoning to all inferential problems, including inferring parameters of statistical models, is sometimes called a Bayesian approach. However, this term tends to elicit very strong reactions (either positive or negative, depending on who you ask), so we prefer the more neutral term "probabilistic approach". Besides, we will often use techniques such as maximum likelihood estimation, which are not Bayesian methods, but certainly fall within the probabilistic paradigm.

Rather than describing a cookbook of different heuristic methods, this book stresses a principled model-based approach to machine learning. For any given model, a variety of algorithms

can often be applied. Conversely, any given algorithm can often be applied to a variety of models. This kind of modularity, where we distinguish model from algorithm, is good pedagogy and good engineering.

We will often use the language of graphical models to specify our models in a concise and intuitive way. In addition to aiding comprehension, the graph structure aids in developing efficient algorithms, as we will see. However, this book is not primarily about graphical models; it is about probabilistic modeling in general.

A practical approach

Nearly all of the methods described in this book have been implemented in a MATLAB software package called PMTK, which stands for probabilistic modeling toolkit. This is freely available from `pmtk3.googlecode.com` (the digit 3 refers to the third edition of the toolkit, which is the one used in this version of the book). There are also a variety of supporting files, written by other people, available at `pmtksupport.googlecode.com`. These will be downloaded automatically, if you follow the setup instructions described on the PMTK website.

MATLAB is a high-level, interactive scripting language ideally suited to numerical computation and data visualization, and can be purchased from `www.mathworks.com`. Some of the code requires the Statistics toolbox, which needs to be purchased separately. There is also a free version of Matlab called Octave, available at `http://www.gnu.org/software/octave/`, which supports most of the functionality of MATLAB. Some (but not all) of the code in this book also works in Octave. See the PMTK website for details.

PMTK was used to generate many of the figures in this book; the source code for these figures is included on the PMTK website, allowing the reader to easily see the effects of changing the data or algorithm or parameter settings. The book refers to files by name, e.g., `naiveBayesFit`. In order to find the corresponding file, you can use two methods: within Matlab you can type `which naiveBayesFit` and it will return the full path to the file; or, if you do not have Matlab but want to read the source code anyway, you can use your favorite search engine, which should return the corresponding file from the `pmtk3.googlecode.com` website.

Details on *how to use* PMTK can be found on the website, which will be udpated over time. Details on the *underlying theory* behind these methods can be found in this book.

Acknowledgments

A book this large is obviously a team effort. I would especially like to thank the following people: my wife Margaret, for keeping the home fires burning as I toiled away in my office for the last six years; Matt Dunham, who created many of the figures in this book, and who wrote much of the code in PMTK; Baback Moghaddam (RIP), who gave extremely detailed feedback on every page of an earlier draft of the book; Chris Williams, who also gave very detailed feedback; Cody Severinski and Wei-Lwun Lu, who assisted with figures; generations of UBC students, who gave helpful comments on earlier drafts; Daphne Koller, Nir Friedman, and Chris Manning, for letting me use their latex style files; Stanford University, Google Research and Skyline College for hosting me during part of my sabbatical; and various Canadian funding agencies (NSERC, CRC and CIFAR) who have supported me financially over the years.

In addition, I would like to thank the following people for giving me helpful feedback on

parts of the book, and/or for sharing figures, code, exercises or even (in some cases) text: David Blei, Sebastien Bratieres, Hannes Bretschneider, Greg Corrado, Arnaud Doucet, Mario Figueiredo, Nando de Freitas, Mark Girolami, Gabriel Goh, Tom Griffiths, Katherine Heller, Geoff Hinton, Aapo Hyvarinen, Tommi Jaakkola, Mike Jordan, Charles Kemp, Emtiyaz Khan, Bonnie Kirkpatrick, Daphne Koller, Zico Kolter, Honglak Lee, Julien Mairal, Andrew McPherson, Tom Minka, Ian Nabney, Robert Piche, Arthur Pope, Carl Rassmussen, Ryan Rifkin, Ruslan Salakhutdinov, Mark Schmidt, Daniel Selsam, David Sontag, Erik Sudderth, Josh Tenenbaum, Martin Wainwright, Yair Weiss, Kai Yu.

Kevin Patrick Murphy
Palo Alto, California
June 2012 (Minor updates February 2013)

Contents