Mathematics for Machine Learning

The fundamental mathematical tools needed to understand machine learning include linear algebra, analytic geometry, matrix decompositions, vector calculus, optimization, probability, and statistics. These topics are traditionally taught in disparate courses, making it hard for data science or computer science students, or professionals, to efficiently learn the mathematics.

This self-contained textbook bridges the gap between mathematical and machine learning texts, introducing the mathematical concepts with a minimum of prerequisites. It uses these concepts to derive four central machine learning methods: linear regression, principal component analysis, Gaussian mixture models, and support vector machines. For students and others with a mathematical background, these derivations provide a starting point to machine learning texts. For those learning the mathematics for the first time, the methods help build intuition and practical experience with applying mathematical concepts.

Every chapter includes worked examples and exercises to test understanding. Programming tutorials are offered on the book's web site.

MARC PETER DEISENROTH is the DeepMind Chair in Artificial Intelligence at University College London. Prior to this, Marc was a faculty member at Imperial College London. His research areas include data-efficient learning, probabilistic modeling, and autonomous decision making. His research received Best Paper Awards at the ICRA 2014 and the ICCAS 2016. Marc has been awarded the President's Award for Outstanding Early Career Researcher at Imperial College London, a Google Faculty Research Award, and a Microsoft PhD grant.

A. ALDO FAISAL leads the Brain & Behaviour Lab at Imperial College London, where he is faculty at the Departments of Bioengineering and Computing and a Fellow of the Data Science Institute. He is the director of the 20Mio \pounds United Kingdom Research and Innovation (UKRI) Center for Doctoral Training in AI for Healthcare. He obtained a PhD in computational neuroscience at Cambridge University and became Junior Research Fellow in the Computational and Biological Learning Lab. His research is at the interface of neuroscience and machine learning to understand and reverse engineer brains and behavior.

CHENG SOON ONG is Principal Research Scientist at the Machine Learning Research Group, Data61, CSIRO and Adjunct Associate Professor at the Australian National University. His research focuses on enabling scientific discovery by extending statistical machine learning methods. He received his PhD in computer science at Australian National University in 2005. He has been a lecturer in the Department of Computer Science at ETH Zürich, and has worked in the Diagnostic Genomics Team at NICTA in Melbourne.

Mathematics for Machine Learning

Marc Peter Deisenroth University College London

> A. Aldo Faisal Imperial College London

> Cheng Soon Ong Data61, CSIRO



CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org Information on this title: www.cambridge.org/9781108470049 DOI: 10.1017/9781108679930

© Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

Printed in Singapore by Markono Print Media Pte Ltd

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data Names: Deisenroth, Marc Peter, author. | Faisal, A. Aldo, author. | Ong, Cheng Soon, author. Title: Mathematics for machine learning / Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong. Description: Cambridge ; New York, NY : Cambridge University Press, 2020. | Includes bibliographical references and index. Identifiers: LCCN 2019040762 (print) | LCCN 2019040763 (ebook) | ISBN 9781108470049 (hardback) | ISBN 9781108455145 (paperback) | ISBN 9781108679930 (epub) Subjects: LCCH: Machine learning–Mathematics. Classification: LCC Q325.5 .D45 2020 (print) | LCC Q325.5 (ebook) | DDC 006.3/1–dc23 LC record available at https://lccn.loc.gov/2019040762

LC ebook record available at https://lccn.loc.gov/2019040763

ISBN 978-1-108-47004-9 Hardback ISBN 978-1-108-45514-5 Paperback

Additional resources for this publication at https://mml-book.com.

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

List of Symbols Preface Acknowledgments		ix xi xv
Part I	Mathematical Foundations	
1	Introduction and Motivation	3
1.1	Finding Words for Intuitions	3
1.2	Two Ways to Read This Book	5
1.3	Exercises and Feedback	7
2	Linear Algebra	8
2.1	Systems of Linear Equations	10
2.2	Matrices	12
2.3	Solving Systems of Linear Equations	17
2.4	Vector Spaces	24
2.5	Linear Independence	29
2.6	Basis and Rank	33
2.7	Linear Mappings	36
2.8	Affine Spaces	48
2.9	Further Reading	50
	Exercises	51
3	Analytic Geometry	57
3.1	Norms	58
3.2	Inner Products	59
3.3	Lengths and Distances	61
3.4	Angles and Orthogonality	63
3.5	Orthonormal Basis	65
3.6	Orthogonal Complement	65
3.7	Inner Product of Functions	66
3.8	Orthogonal Projections	67
3.9	Rotations	76
3.10	Further Reading	79
	Exercises	80

v

4.1Determinant and Trace834.2Eigenvalues and Eigenvectors884.3Cholesky Decomposition964.4Eigendecomposition and Diagonalization984.5Singular Value Decomposition1014.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	vi		Contents
4.2Eigenvalues and Eigenvectors884.3Cholesky Decomposition964.4Eigendecomposition and Diagonalization984.5Singular Value Decomposition1014.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4	Matrix Decompositions	82
4.3Cholesky Decomposition964.4Eigendecomposition and Diagonalization984.5Singular Value Decomposition1014.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.1	Determinant and Trace	83
4.4Eigendecomposition and Diagonalization984.5Singular Value Decomposition1014.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.2	Eigenvalues and Eigenvectors	88
4.5Singular Value Decomposition1014.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.3	Cholesky Decomposition	96
4.6Matrix Approximation1114.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.4	Eigendecomposition and Diagonalization	98
4.7Matrix Phylogeny1154.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.5	Singular Value Decomposition	101
4.8Further Reading116Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.6	Matrix Approximation	111
Exercises1185Vector Calculus1205.1Differentiation of Univariate Functions122	4.7	Matrix Phylogeny	115
5Vector Calculus1205.1Differentiation of Univariate Functions122	4.8	Further Reading	116
5.1Differentiation of Univariate Functions122		Exercises	118
	5	Vector Calculus	120
	5.1	Differentiation of Univariate Functions	122
5.2 Partial Differentiation and Gradients 126	5.2	Partial Differentiation and Gradients	126
5.3 Gradients of Vector-Valued Functions 129	5.3	Gradients of Vector-Valued Functions	129
5.4 Gradients of Matrices 135	5.4	Gradients of Matrices	135
5.5 Useful Identities for Computing Gradients 138	5.5	Useful Identities for Computing Gradients	138
5.6 Backpropagation and Automatic Differentiation 138	5.6	Backpropagation and Automatic Differentiation	138
5.7 Higher-Order Derivatives 143	5.7	Higher-Order Derivatives	143
5.8 Linearization and Multivariate Taylor Series 144	5.8	Linearization and Multivariate Taylor Series	144
5.9 Further Reading 149	5.9	Further Reading	149
Exercises 150		Exercises	150
6 Probability and Distributions 152	6	Probability and Distributions	152
6.1 Construction of a Probability Space 152	6.1	Construction of a Probability Space	152
6.2 Discrete and Continuous Probabilities 157	6.2	Discrete and Continuous Probabilities	157
6.3 Sum Rule, Product Rule, and Bayes' Theorem 163	6.3	Sum Rule, Product Rule, and Bayes' Theorem	163
6.4 Summary Statistics and Independence 165	6.4	Summary Statistics and Independence	165
6.5 Gaussian Distribution 175	6.5	Gaussian Distribution	175
6.6 Conjugacy and the Exponential Family 182	6.6	Conjugacy and the Exponential Family	182
6.7 Change of Variables/Inverse Transform 191	6.7	Change of Variables/Inverse Transform	191
6.8Further Reading197	6.8	Further Reading	197
Exercises 198		Exercises	198
7 Continuous Optimization 201	7	Continuous Optimization	201
7.1 Optimization Using Gradient Descent 203	7.1	Optimization Using Gradient Descent	203
7.2 Constrained Optimization and Lagrange Multipliers 208	7.2	Constrained Optimization and Lagrange Multipliers	208
7.3 Convex Optimization 211	7.3	Convex Optimization	211
7.4Further Reading220	7.4	Further Reading	220
Exercises 221		Exercises	221
Part II Central Machine Learning Problems	Part II	Central Machine Learning Problems	
8 When Models Meet Data 225	8	When Models Meet Data	225
8.1 Data, Models, and Learning 225	8.1	Data, Models, and Learning	225
8.2 Empirical Risk Minimization 232	8.2	Empirical Risk Minimization	232

Contents		vii
8.3	Parameter Estimation	238
8.4	Probabilistic Modeling and Inference	244
8.5	Directed Graphical Models	249
8.6	Model Selection	254
9	Linear Regression	260
9.1	Problem Formulation	261
9.2	Parameter Estimation	263
9.3	Bayesian Linear Regression	273
9.4	Maximum Likelihood as Orthogonal Projection	282
9.5	Further Reading	283
10	Dimensionality Reduction with Principal Component Analysis	286
10.1	Problem Setting	286
10.2	Maximum Variance Perspective	289
10.3	Projection Perspective	293
10.4	Eigenvector Computation and Low-Rank Approximations	300
10.5	PCA in High Dimensions	302
10.6	Key Steps of PCA in Practice	303
10.7	Latent Variable Perspective	306
10.8	Further Reading	310
11	Density Estimation with Gaussian Mixture Models	314
11.1	Gaussian Mixture Model	315
11.2	Parameter Learning via Maximum Likelihood	316
11.3	EM Algorithm	325
11.4	Latent-Variable Perspective	328
11.5	Further Reading	332
12	Classification with Support Vector Machines	335
12.1	Separating Hyperplanes	337
12.2	Primal Support Vector Machine	338
12.3	Dual Support Vector Machine	347
12.4	Kernels	351
12.5	Numerical Solution	353
12.6	Further Reading	355
Refere	nces	357
Index		367

List of Symbols

Symbol	Typical meaning
$\overline{a, b, c, \alpha, \beta, \gamma}$	Scalars are lowercase
$oldsymbol{x},oldsymbol{y},oldsymbol{z}$	Vectors are bold lowercase
A, B, C	Matrices are bold uppercase
$oldsymbol{x}^ op,oldsymbol{A}^ op$	Transpose of a vector or matrix
A^{-1}	Inverse of a matrix
$\langle oldsymbol{x},oldsymbol{y} angle$	Inner product of x and y
$\mathbf{x}^{\top}\mathbf{y}$	Dot product of x and y
$B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$	(Ordered) tuple
$B = [b_1, b_2, b_3]$	Matrix of column vectors stacked horizontally
$\mathcal{B} = \{oldsymbol{b}_1, oldsymbol{b}_2, oldsymbol{b}_3\}$	Set of vectors (unordered)
\mathbb{Z},\mathbb{N}	Integers and natural numbers, respectively
\mathbb{R}, \mathbb{C}	Real and complex numbers, respectively
\mathbb{R}^{n}	n-dimensional vector space of real numbers
$\forall x$	Universal quantifier: for all x
$\exists x$	Existential quantifier: there exists x
a := b	a is defined as b
a =: b	b is defined as a
$a \propto b$	a is proportional to b, i.e., $a = \text{constant} \cdot b$
$g \circ f$	Function composition: " g after f "
\iff	If and only if
\implies	Implies
\mathcal{A},\mathcal{C}	Sets
$a \in \mathcal{A}$	a is an element of the set \mathcal{A}
Ø	Empty set
D	Number of dimensions; indexed by $d = 1, \dots, D$
N	Number of data points; indexed by $n = 1, \ldots, N$
I_m	Identity matrix of size $m \times m$
$0_{m,n}$	Matrix of zeros of size $m \times n$
$1_{m,n}$	Matrix of ones of size $m \times n$
$oldsymbol{e}_i$	Standard/canonical vector (where i is the
	component that is 1)
dim(V)	Dimensionality of vector space V

CAMBRIDGE

Cambridge University Press 978-1-108-47004-9 — Mathematics for Machine Learning Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong Frontmatter <u>More Information</u>

X	List of Symbols
Symbol	Typical meaning
rk(A)	Rank of matrix <i>A</i>
$\operatorname{Im}(\Phi)$	Image of linear mapping Φ
$\ker(\Phi)$	Kernel (null space) of a linear mapping Φ
$\operatorname{span}[{m b}_1]$	Span (generating set) of b_1
tr(A)	Trace of <i>A</i>
$\det(\boldsymbol{A})$	Determinant of A
$ \cdot $	Absolute value or determinant (depending on context)
$\ \cdot\ $	Norm; Euclidean unless specified
λ	Eigenvalue or Lagrange multiplier
E_{λ}	Eigenspace corresponding to eigenvalue λ
θ	Parameter vector
$rac{\partial f}{\partial x}$	Partial derivative of f with respect to x
$\frac{\mathrm{d}f}{\mathrm{d}x}$	Total derivative of f with respect to x
$egin{array}{c} abla & \mathcal{L} \\ abla & \mathcal{L} \end{array}$	Gradient
\mathfrak{L}	Lagrangian
\mathcal{L}	Negative log-likelihood
$\binom{n}{k}$	Binomial coefficient, n choose k
$\mathbb{V}_X[oldsymbol{x}]$	Variance of \boldsymbol{x} with respect to the random variable X
$\mathbb{E}_X[oldsymbol{x}]$	Expectation of \boldsymbol{x} with respect to the random variable X
$\mathrm{Cov}_{X,Y}[oldsymbol{x},oldsymbol{y}]$	Covariance between x and y .
$X \perp\!\!\!\perp Y Z$	X is conditionally independent of Y given Z
$X \sim p$	Random variable X is distributed according to p
$\mathcal{N}oldsymbol{(\mu, \Sigma)}$	Gaussian distribution with mean μ and covariance Σ
$\operatorname{Ber}(\mu)$	Bernoulli distribution with parameter μ
$\operatorname{Bin}(N,\mu)$	Binomial distribution with parameters N, μ
$Beta(\alpha,\beta)$	Beta distribution with parameters α, β

List of Abbreviations and Acronyms

Acronym	Meaning
e.g.	Exempli gratia (Latin: for example)
GMM	Gaussian mixture model
i.e.	Id est (Latin: this means)
i.i.d.	Independent, identically distributed
MAP	Maximum a posteriori
MLE	Maximum likelihood estimation/estimator
ONB	Orthonormal basis
PCA	Principal component analysis
PPCA	Probabilistic principal component analysis
REF	Row-echelon form
SPD	Symmetric, positive definite
SVM	Support vector machine

Preface

Machine learning is the latest in a long line of attempts to distill human knowledge and reasoning into a form that is suitable for constructing machines and engineering automated systems. As machine learning becomes more ubiquitous and its software packages become easier to use, it is natural and desirable that the low-level technical details are abstracted away and hidden from the practitioner. However, this brings with it the danger that a practitioner becomes unaware of the design decisions and, hence, the limits of machine learning algorithms.

The enthusiastic practitioner who is interested to learn more about the magic behind successful machine learning algorithms currently faces a daunting set of prerequisite knowledge:

- Programming languages and data analysis tools
- Large-scale computation and the associated frameworks
- Mathematics and statistics and how machine learning builds on it

At universities, introductory courses on machine learning tend to spend early parts of the course covering some of these prerequisites. For historical reasons, courses in machine learning tend to be taught in the computer science department, where students are often trained in the first two areas of knowledge, but not so much in mathematics and statistics.

Current machine learning textbooks primarily focus on machine learning algorithms and methodologies and assume that the reader is competent in mathematics and statistics. Therefore, these books only spend one or two chapters of background mathematics, either at the beginning of the book or as appendices. We have found many people who want to delve into the foundations of basic machine learning methods who struggle with the mathematical knowledge required to read a machine learning textbook. Having taught undergraduate and graduate courses at universities, we find that the gap between high school mathematics and the mathematics level required to read a standard machine learning textbook is too big for many people.

This book brings the mathematical foundations of basic machine learning concepts to the fore and collects the information in a single place so that this skills gap is narrowed or even closed.

xii

Preface

Why Another Book on Machine Learning?

Machine learning builds upon the language of mathematics to express concepts that seem intuitively obvious but that are surprisingly difficult to formalize. Once formalized properly, we can gain insights into the task we want to solve. One common complaint of students of mathematics around the globe is that the topics covered seem to have little relevance to practical problems. We believe that machine learning is an obvious and direct motivation for people to learn mathematics.

This book is intended to be a guidebook to the vast mathematical literature that forms the foundations of modern machine learning. We motivate the need for mathematical concepts by directly pointing out their usefulness in the context of fundamental machine learning problems. In the interest of keeping the book short, many details and more advanced concepts have been left out. Equipped with the basic concepts presented here, and how they fit into the larger context of machine learning, the reader can find numerous resources for further study, which we provide at the end of the respective chapters. For readers with a mathematical background, this book provides a brief but precisely stated glimpse of machine learning. In contrast to other books that focus on methods and models of machine learning (MacKay, 2003; Bishop, 2006; Alpaydin, 2010; Murphy, 2012; Barber, 2012; Shalev-Shwartz and Ben-David, 2014; Rogers and Girolami, 2016) or programmatic aspects of machine learning (Müller and Guido, 2016; Raschka and Mirjalili, 2017; Chollet and Allaire, 2018), we provide only four representative examples of machine learning algorithms. Instead, we focus on the mathematical concepts behind the models themselves. We hope that readers will be able to gain a deeper understanding of the basic questions in machine learning and connect practical questions arising from the use of machine learning with fundamental choices in the mathematical model.

We do not aim to write a classical machine learning book. Instead, our intention is to provide the mathematical background, applied to four central machine learning problems, to make it easier to read other machine learning textbooks.

Who Is the Target Audience?

As applications of machine learning become widespread in society, we believe that everybody should have some understanding of its underlying principles. This book is written in an academic mathematical style, which enables us to be precise about the concepts behind machine learning. We encourage readers unfamiliar with this seemingly terse style to persevere and to keep the goals of each topic in mind. We sprinkle comments and remarks throughout the text, in the hope that it provides useful guidance with respect to the big picture.

The book assumes the reader to have mathematical knowledge commonly covered in high school mathematics and physics. For example, the reader should have seen derivatives and integrals before, and geometric vectors in two or three dimensions. Starting from there, we generalize these concepts. Therefore, the

"Math is linked in the popular mind with phobia and anxiety. You'd think we're discussing spiders." (Strogatz, 2014, 281)

Preface

target audience of the book includes undergraduate university students, evening learners and learners participating in online machine learning courses.

In analogy to music, there are three types of interaction that people have with machine learning:

Astute Listener The democratization of machine learning by the provision of open-source software, online tutorials and cloud-based tools allows users to not worry about the specifics of pipelines. Users can focus on extracting insights from data using off-the-shelf tools. This enables non-tech-savvy domain experts to benefit from machine learning. This is similar to listening to music; the user is able to choose and discern between different types of machine learning, and benefits from it. More experienced users are like music critics, asking important questions about the application of machine learning in society such as ethics, fairness and privacy of the individual. We hope that this book provides a foundation for thinking about the certification and risk management of machine learning systems and allows them to use their domain expertise to build better machine learning systems.

Experienced Artist Skilled practitioners of machine learning can plug and play different tools and libraries into an analysis pipeline. The stereotypical practitioner would be a data scientist or engineer who understands machine learning interfaces and their use cases and is able to perform wonderful feats of prediction from data. This is similar to a virtuoso playing music, where highly skilled practitioners can bring existing instruments to life and bring enjoyment to their audience. Using the mathematics presented here as a primer, practitioners would be able to understand the benefits and limits of their favourite method, and to extend and generalize existing machine learning algorithms. We hope that this book provides the impetus for more rigorous and principled development of machine learning methods.

Fledgling Composer As machine learning is applied to new domains, developers of machine learning need to develop new methods and extend existing algorithms. They are often researchers who need to understand the mathematical basis of machine learning and uncover relationships between different tasks. This is similar to composers of music who, within the rules and structure of musical theory, create new and amazing pieces. We hope this book provides a high-level overview of other technical books for people who want to become composers of machine learning. There is a great need in society for new researchers who are able to propose and explore novel approaches for attacking the many challenges of learning from data.

xiii

Acknowledgments

We are grateful to many people who looked at early drafts of the book and suffered through painful expositions of concepts. We tried to implement their ideas that we did not vehemently disagree with. We would like to especially acknowledge Christfried Webers for his careful reading of many parts of the book, and his detailed suggestions on structure and presentation. Many friends and colleagues have also been kind enough to provide their time and energy on different versions of each chapter. We have been lucky to benefit from the generosity of the online community, who have suggested improvements via github.com, which greatly improved the book.

The following people have found bugs, proposed clarifications and suggested relevant literature, either via github.com or personal communication. Their names are sorted alphabetically.

Abdul-Ganiy Usman Adam Gaier Adele Jackson Aditya Menon Alasdair Tran Aleksandar Krnjaic Alexander Makrigiorgos Alfredo Canziani Ali Shafti Amr Khalifa Andrew Tanggara Angus Gruen Antal A. Buss Antoine Toisoul Le Cann Areg Sarvazyan Artem Artemev Artyom Stepanov Bill Kromydas **Bob Williamson Boon Ping Lim** Chao Qu Cheng Li Chris Sherlock

Christopher Gray Daniel McNamara Daniel Wood Darren Siegel David Johnston Dawei Chen Ellen Broad Fengkuangtian Zhu Fiona Condon Georgios Theodorou He Xin Irene Raissa Kameni Jakub Nabaglo James Hensman Jamie Liu Jean Kaddour Jean-Paul Ebejer Jerry Qiang Jitesh Sindhare John Lloyd Jonas Ngnawe Jon Martin Justin Hsi

XV

CAMBRIDGE

Cambridge University Press 978-1-108-47004-9 — Mathematics for Machine Learning Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong Frontmatter <u>More Information</u>

xvi

Acknowledgments

Kai Arulkumaran Kamil Dreczkowski Lily Wang Lionel Tondji Ngoupeyou Lydia Knüfing Mahmoud Aslan Mark Hartenstein Mark van der Wilk Markus Hegland Martin Hewing Matthew Alger Matthew Lee Maximus McCann Mengyan Zhang Michael Bennett Michael Pedersen Minjeong Shin Mohammad Malekzadeh Naveen Kumar Nico Montali Oscar Armas Patrick Henriksen Patrick Wieschollek Pattarawat Chormai Paul Kelly Petros Christodoulou Piotr Januszewski Pranav Subramani Quyu Kong Ragib Zaman Rui Zhang **Ryan-Rhys Griffiths** Salomon Kabongo Samuel Ogunmola

Sandeep Mavadia Sarvesh Nikumbh Sebastian Raschka Senanayak Sesh Kumar Karri Seung-Heon Baek Shahbaz Chaudhary Shakir Mohamed Shawn Berry Sheikh Abdul Raheem Ali Sheng Xue Sridhar Thiagarajan Syed Nouman Hasany Szymon Brych Thomas Bühler Timur Sharapov Tom Melamed Vincent Adam Vincent Dutordoir Vu Minh Wasim Aftab Wen Zhi Wojciech Stokowiec Xiaonan Chong Xiaowei Zhang Yazhou Hao Yicheng Luo Young Lee Yu Lu Yun Cheng Yuxiao Huang Zac Cranko Zijian Cao Zoe Nolan

Contributors through github, whose real names were not listed on their github profile, are the following:

SamDataMad	insad	empet
bumptiousmonkey	HorizonP	victorBigand
idoamihai	cs-maillist	17SKYE
deepakiim	kudo23	jessjing1995

We are also very grateful to Parameswaran Raman and the many anonymous reviewers, organized by Cambridge University Press, who read one or more

Acknowledgments

xvii

chapters of earlier versions of the manuscript, and provided constructive criticism that led to considerable improvements. A special mention goes to Dinesh Singh Negi, our LATEX support for detailed and prompt advice about LATEX-related issues. Last but not least, we are very grateful to our editor Lauren Cowles, who has been patiently guiding us through the gestation process of this book.