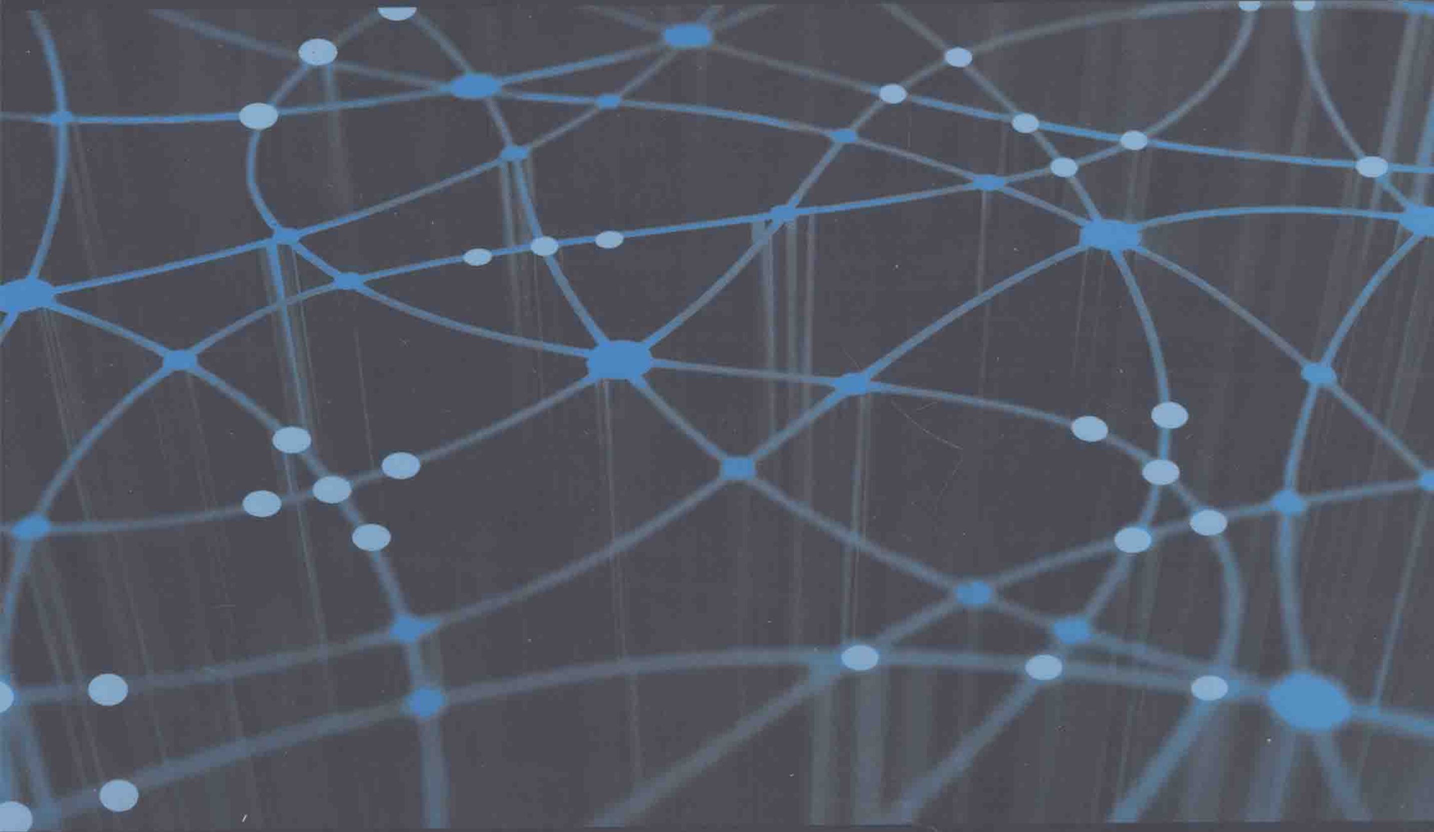
PROBABILISTIC GRAPHICAL MODELS

PRINCIPLES AND TECHNIQUES



DAPHNE KOLLER AND NIR FRIEDMAN

Probabilistic Graphical Models

Principles and Techniques

常州大学图书馆
藏书章

Daphne Koller

Nir Friedman

The MIT Press
Cambridge, Massachusetts
London, England

©2009 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

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

This book was set by the authors in LaTeX2$_\epsilon$.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Koller, Daphne.
Probabilistic Graphical Models: Principles and Techniques / Daphne Koller and Nir Friedman.
p. cm. – (Adaptive computation and machine learning)
Includes bibliographical references and index.
ISBN 978-0-262-01319-2 (hardcover : alk. paper)
1. Graphical modeling (Statistics) 2. Bayesian statistical decision theory—Graphic methods. I. Koller, Daphne. II. Friedman, Nir.
QA279.5.K65 2010
519.5'420285–dc22
2009008615

10 9 8 7 6 5 4

Probabilistic Graphical Models

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, 2nd ed., Peter Spirtes, Clark Glymour, and Richard Scheines

Principles of Data Mining, David Hand, Heikki Mannila, and Padhraic Smyth

Bioinformatics: The Machine Learning Approach, 2nd ed., 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. Grünwald

Introduction to Statistical Relational Learning, Lise Getoor and Ben Taskar, eds.

Probabilistic Graphical Models: Principles and Techniques, Daphne Koller and Nir Friedman

此为试读，需要完整PDF请访问：www.ertongbook.com

To our families

my parents Dov and Ditza
my husband Dan
my daughters Natalie and Maya
D.K.

my parents Noga and Gad
my wife Yael
my children Roy and Lior
N.F.

As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality.

Albert Einstein, 1956

When we try to pick out anything by itself, we find that it is bound fast by a thousand invisible cords that cannot be broken, to everything in the universe.

John Muir, 1869

The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful . . . Therefore the true logic for this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

James Clerk Maxwell, 1850

The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which ofttimes they are unable to account.

Pierre Simon Laplace, 1819

Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

Stephen Jay Gould

Acknowledgments

This book owes a considerable debt of gratitude to the many people who contributed to its creation, and to those who have influenced our work and our thinking over the years.

First and foremost, we want to thank our students, who, by asking the right questions, and forcing us to formulate clear and precise answers, were directly responsible for the inception of this book and for any clarity of presentation.

We have been fortunate to share the same mentors, who have had a significant impact on our development as researchers and as teachers: Joe Halpern, Stuart Russell. Much of our core views on probabilistic models have been influenced by Judea Pearl. Judea through his persuasive writing and vivid presentations inspired us, and many other researchers of our generation, to plunge into research in this field.

There are many people whose conversations with us have helped us in thinking through some of the more difficult concepts in the book: Nando de Freitas, Gal Elidan, Dan Geiger, Amir Globerson, Uri Lerner, Chris Meek, David Sontag, Yair Weiss, and Ramin Zabih. Others, in conversations and collaborations over the year, have also influenced our thinking and the presentation of the material: Pieter Abbeel, Jeff Bilmes, Craig Boutilier, Moises Goldszmidt, Carlos Guestrin, David Heckerman, Eric Horvitz, Tommi Jaakkola, Michael Jordan, Kevin Murphy, Andrew Ng, Ben Taskar, and Sebastian Thrun.

We especially want to acknowledge Gal Elidan for constant encouragement, valuable feedback, and logistic support at many critical junctions, throughout the long years of writing this book.

Over the course of the years of work on this book, many people have contributed to it by providing insights, engaging in enlightening discussions, and giving valuable feedback. It is impossible to individually acknowledge all of the people who made such contributions. However, we specifically wish to express our gratitude to those people who read large parts of the book and gave detailed feedback: Rahul Biswas, James Cussens, James Diebel, Yoni Donner, Tal El-Hay, Gal Elidan, Stanislav Funiak, Amir Globerson, Russ Greiner, Carlos Guestrin, Tim Heilman, Geremy Heitz, Maureen Hillenmeyer, Ariel Jaimovich, Tommy Kaplan, Jonathan Laserson, Ken Levine, Brian Milch, Kevin Murphy, Ben Packer, Ronald Parr, Dana Pe'er, and Christian Shelton.

We are deeply grateful to the following people, who contributed specific text and/or figures, mostly to the case studies and concept boxes without which this book would be far less interesting: Gal Elidan, to chapter 11, chapter 18, and chapter 19; Stephen Gould, to chapter 4 and chapter 13; Vladimir Jojic, to chapter 12; Jonathan Laserson, to chapter 19; Uri Lerner, to chapter 14; Andrew McCallum and Charles Sutton, to chapter 4; Brian Milch, to chapter 6; Kevin

Murphy, to chapter 15; and Benjamin Packer, to many of the exercises used throughout the book. In addition, we are very grateful to Amir Globerson, David Sontag and Yair Weiss whose insights on chapter 13 played a key role in the development of the material in that chapter.

Special thanks are due to Bob Prior at MIT Press who convinced us to go ahead with this project and was constantly supportive, enthusiastic and patient in the face of the recurring delays and missed deadlines. We thank Greg McNamee, our copy editor, and Mary Reilly, our artist, for their help in improving this book considerably. We thank Chris Manning, for allowing us to use his LaTeX macros for typesetting this book, and for providing useful advice on how to use them. And we thank Miles Davis for invaluable technical support.

We also wish to thank the many colleagues who used drafts of this book in teaching provided enthusiastic feedback that encouraged us to continue this project at times where it seemed unending. Sebastian Thrun deserves a special note of thanks, for forcing us to set a deadline for completion of this book and to stick to it.

We also want to thank the past and present members of the DAGS group at Stanford, and the Computational Biology group at the Hebrew University, many of whom also contributed ideas, insights, and useful comments. We specifically want to thank them for bearing with us while we devoted far too much of our time to working on this book.

Finally, noone deserves our thanks more than our long-suffering families — Natalie Anna Koller Avida, Maya Rika Koller Avida, and Dan Avida; Lior, Roy, and Yael Friedman — for their continued love, support, and patience, as they watched us work evenings and weekends to complete this book. We could never have done this without you.

Contents

Acknowledgments **xxiii**

List of Figures **xxv**

List of Algorithms **xxxi**

List of Boxes **xxxiii**

1 Introduction **1**

1.1 Motivation 1
1.2 Structured Probabilistic Models 2
1.2.1 Probabilistic Graphical Models 3
1.2.2 Representation, Inference, Learning 5
1.3 Overview and Roadmap 6
1.3.1 Overview of Chapters 6
1.3.2 Reader's Guide 9
1.3.3 Connection to Other Disciplines 11
1.4 Historical Notes 12

2 Foundations **15**

2.1 Probability Theory 15
2.1.1 Probability Distributions 15
2.1.2 Basic Concepts in Probability 18
2.1.3 Random Variables and Joint Distributions 19
2.1.4 Independence and Conditional Independence 23
2.1.5 Querying a Distribution 25
2.1.6 Continuous Spaces 27
2.1.7 Expectation and Variance 31
2.2 Graphs 34
2.2.1 Nodes and Edges 34
2.2.2 Subgraphs 35
2.2.3 Paths and Trails 36

2.2.4 Cycles and Loops 36
2.3 Relevant Literature 39
2.4 Exercises 39

I Representation 43

***3 The Bayesian Network Representation* 45**
3.1 Exploiting Independence Properties 45
3.1.1 Independent Random Variables 45
3.1.2 The Conditional Parameterization 46
3.1.3 The Naive Bayes Model 48
3.2 Bayesian Networks 51
3.2.1 The Student Example Revisited 52
3.2.2 Basic Independencies in Bayesian Networks 56
3.2.3 Graphs and Distributions 60
3.3 Independencies in Graphs 68
3.3.1 D-separation 69
3.3.2 Soundness and Completeness 72
3.3.3 An Algorithm for d-Separation 74
3.3.4 I-Equivalence 76
3.4 From Distributions to Graphs 78
3.4.1 Minimal I-Maps 78
3.4.2 Perfect Maps 81
3.4.3 Finding Perfect Maps ⋆ 83
3.5 Summary 92
3.6 Relevant Literature 93
3.7 Exercises 96

***4 Undirected Graphical Models* 103**
4.1 The Misconception Example 103
4.2 Parameterization 106
4.2.1 Factors 106
4.2.2 Gibbs Distributions and Markov Networks 108
4.2.3 Reduced Markov Networks 110
4.3 Markov Network Independencies 114
4.3.1 Basic Independencies 114
4.3.2 Independencies Revisited 117
4.3.3 From Distributions to Graphs 120
4.4 Parameterization Revisited 122
4.4.1 Finer-Grained Parameterization 123
4.4.2 Overparameterization 128
4.5 Bayesian Networks and Markov Networks 134
4.5.1 From Bayesian Networks to Markov Networks 134
4.5.2 From Markov Networks to Bayesian Networks 137

此为试读，需要完整PDF请访问：www.ertongbook.com

4.5.3 Chordal Graphs 139
4.6 Partially Directed Models 142
4.6.1 Conditional Random Fields 142
4.6.2 Chain Graph Models ⋆ 148
4.7 Summary and Discussion 151
4.8 Relevant Literature 152
4.9 Exercises 153

***5 Local Probabilistic Models* 157**

5.1 Tabular CPDs 157
5.2 Deterministic CPDs 158
5.2.1 Representation 158
5.2.2 Independencies 159
5.3 Context-Specific CPDs 162
5.3.1 Representation 162
5.3.2 Independencies 171
5.4 Independence of Causal Influence 175
5.4.1 The Noisy-Or Model 175
5.4.2 Generalized Linear Models 178
5.4.3 The General Formulation 182
5.4.4 Independencies 184
5.5 Continuous Variables 185
5.5.1 Hybrid Models 189
5.6 Conditional Bayesian Networks 191
5.7 Summary 193
5.8 Relevant Literature 194
5.9 Exercises 195

***6 Template-Based Representations* 199**

6.1 Introduction 199
6.2 Temporal Models 200
6.2.1 Basic Assumptions 201
6.2.2 Dynamic Bayesian Networks 202
6.2.3 State-Observation Models 207
6.3 Template Variables and Template Factors 212
6.4 Directed Probabilistic Models for Object-Relational Domains 216
6.4.1 Plate Models 216
6.4.2 Probabilistic Relational Models 222
6.5 Undirected Representation 228
6.6 Structural Uncertainty ⋆ 232
6.6.1 Relational Uncertainty 233
6.6.2 Object Uncertainty 235
6.7 Summary 240
6.8 Relevant Literature 242
6.9 Exercises 243

7 Gaussian Network Models **247**
7.1 Multivariate Gaussians 247
7.1.1 Basic Parameterization 247
7.1.2 Operations on Gaussians 249
7.1.3 Independencies in Gaussians 250
7.2 Gaussian Bayesian Networks 251
7.3 Gaussian Markov Random Fields 254
7.4 Summary 257
7.5 Relevant Literature 258
7.6 Exercises 258

8 The Exponential Family **261**
8.1 Introduction 261
8.2 Exponential Families 261
8.2.1 Linear Exponential Families 263
8.3 Factored Exponential Families 266
8.3.1 Product Distributions 266
8.3.2 Bayesian Networks 267
8.4 Entropy and Relative Entropy 269
8.4.1 Entropy 269
8.4.2 Relative Entropy 272
8.5 Projections 273
8.5.1 Comparison 274
8.5.2 M-Projections 277
8.5.3 I-Projections 282
8.6 Summary 282
8.7 Relevant Literature 283
8.8 Exercises 283

II Inference 285

9 Variable Elimination **287**
9.1 Analysis of Complexity 288
9.1.1 Analysis of Exact Inference 288
9.1.2 Analysis of Approximate Inference 290
9.2 Variable Elimination: The Basic Ideas 292
9.3 Variable Elimination 296
9.3.1 Basic Elimination 297
9.3.2 Dealing with Evidence 303
9.4 Complexity and Graph Structure: Variable Elimination 306
9.4.1 Simple Analysis 306
9.4.2 Graph-Theoretic Analysis 306
9.4.3 Finding Elimination Orderings ⋆ 310
9.5 Conditioning ⋆ 315

9.5.1 The Conditioning Algorithm 315
9.5.2 Conditioning and Variable Elimination 318
9.5.3 Graph-Theoretic Analysis 322
9.5.4 Improved Conditioning 323
9.6 Inference with Structured CPDs ★ 325
9.6.1 Independence of Causal Influence 325
9.6.2 Context-Specific Independence 329
9.6.3 Discussion 335
9.7 Summary and Discussion 336
9.8 Relevant Literature 337
9.9 Exercises 338

10 Clique Trees **345**

10.1 Variable Elimination and Clique Trees 345
10.1.1 Cluster Graphs 346
10.1.2 Clique Trees 346
10.2 Message Passing: Sum Product 348
10.2.1 Variable Elimination in a Clique Tree 349
10.2.2 Clique Tree Calibration 355
10.2.3 A Calibrated Clique Tree as a Distribution 361
10.3 Message Passing: Belief Update 364
10.3.1 Message Passing with Division 364
10.3.2 Equivalence of Sum-Product and Belief Update Messages 368
10.3.3 Answering Queries 369
10.4 Constructing a Clique Tree 372
10.4.1 Clique Trees from Variable Elimination 372
10.4.2 Clique Trees from Chordal Graphs 374
10.5 Summary 376
10.6 Relevant Literature 377
10.7 Exercises 378

11 Inference as Optimization **381**

11.1 Introduction 381
11.1.1 Exact Inference Revisited ★ 382
11.1.2 The Energy Functional 384
11.1.3 Optimizing the Energy Functional 386
11.2 Exact Inference as Optimization 386
11.2.1 Fixed-Point Characterization 388
11.2.2 Inference as Optimization 390
11.3 Propagation-Based Approximation 391
11.3.1 A Simple Example 391
11.3.2 Cluster-Graph Belief Propagation 396
11.3.3 Properties of Cluster-Graph Belief Propagation 399
11.3.4 Analyzing Convergence ★ 401
11.3.5 Constructing Cluster Graphs 404

11.3.6 Variational Analysis 411
11.3.7 Other Entropy Approximations ⋆ 414
11.3.8 Discussion 428
11.4 Propagation with Approximate Messages ⋆ 430
11.4.1 Factorized Messages 431
11.4.2 Approximate Message Computation 433
11.4.3 Inference with Approximate Messages 436
11.4.4 Expectation Propagation 442
11.4.5 Variational Analysis 445
11.4.6 Discussion 448
11.5 Structured Variational Approximations 448
11.5.1 The Mean Field Approximation 449
11.5.2 Structured Approximations 456
11.5.3 Local Variational Methods ⋆ 469
11.6 Summary and Discussion 473
11.7 Relevant Literature 475
11.8 Exercises 477

***12 Particle-Based Approximate Inference* 487**

12.1 Forward Sampling 488
12.1.1 Sampling from a Bayesian Network 488
12.1.2 Analysis of Error 490
12.1.3 Conditional Probability Queries 491
12.2 Likelihood Weighting and Importance Sampling 492
12.2.1 Likelihood Weighting: Intuition 492
12.2.2 Importance Sampling 494
12.2.3 Importance Sampling for Bayesian Networks 498
12.2.4 Importance Sampling Revisited 504
12.3 Markov Chain Monte Carlo Methods 505
12.3.1 Gibbs Sampling Algorithm 505
12.3.2 Markov Chains 507
12.3.3 Gibbs Sampling Revisited 512
12.3.4 A Broader Class of Markov Chains ⋆ 515
12.3.5 Using a Markov Chain 518
12.4 Collapsed Particles 526
12.4.1 Collapsed Likelihood Weighting ⋆ 527
12.4.2 Collapsed MCMC 531
12.5 Deterministic Search Methods ⋆ 536
12.6 Summary 540
12.7 Relevant Literature 541
12.8 Exercises 544

***13 MAP Inference* 551**

13.1 Overview 551
13.1.1 Computational Complexity 551

13.1.2 Overview of Solution Methods 552
13.2 Variable Elimination for (Marginal) MAP 554
13.2.1 Max-Product Variable Elimination 554
13.2.2 Finding the Most Probable Assignment 556
13.2.3 Variable Elimination for Marginal MAP ⋆ 559
13.3 Max-Product in Clique Trees 562
13.3.1 Computing Max-Marginals 562
13.3.2 Message Passing as Reparameterization 564
13.3.3 Decoding Max-Marginals 565
13.4 Max-Product Belief Propagation in Loopy Cluster Graphs 567
13.4.1 Standard Max-Product Message Passing 567
13.4.2 Max-Product BP with Counting Numbers ⋆ 572
13.4.3 Discussion 575
13.5 MAP as a Linear Optimization Problem ⋆ 577
13.5.1 The Integer Program Formulation 577
13.5.2 Linear Programming Relaxation 579
13.5.3 Low-Temperature Limits 581
13.6 Using Graph Cuts for MAP 588
13.6.1 Inference Using Graph Cuts 588
13.6.2 Nonbinary Variables 592
13.7 Local Search Algorithms ⋆ 595
13.8 Summary 597
13.9 Relevant Literature 598
13.10 Exercises 601

***14 Inference in Hybrid Networks* 605**

14.1 Introduction 605
14.1.1 Challenges 605
14.1.2 Discretization 606
14.1.3 Overview 607
14.2 Variable Elimination in Gaussian Networks 608
14.2.1 Canonical Forms 609
14.2.2 Sum-Product Algorithms 611
14.2.3 Gaussian Belief Propagation 612
14.3 Hybrid Networks 615
14.3.1 The Difficulties 615
14.3.2 Factor Operations for Hybrid Gaussian Networks 618
14.3.3 EP for CLG Networks 621
14.3.4 An “Exact” CLG Algorithm ⋆ 626
14.4 Nonlinear Dependencies 630
14.4.1 Linearization 631
14.4.2 Expectation Propagation with Gaussian Approximation 637
14.5 Particle-Based Approximation Methods 642
14.5.1 Sampling in Continuous Spaces 642
14.5.2 Forward Sampling in Bayesian Networks 643

14.5.3 MCMC Methods 644
14.5.4 Collapsed Particles 645
14.5.5 Nonparametric Message Passing 646
14.6 Summary and Discussion 646
14.7 Relevant Literature 647
14.8 Exercises 649

15 Inference in Temporal Models **651**
15.1 Inference Tasks 652
15.2 Exact Inference 653
15.2.1 Filtering in State-Observation Models 653
15.2.2 Filtering as Clique Tree Propagation 654
15.2.3 Clique Tree Inference in DBNs 655
15.2.4 Entanglement 656
15.3 Approximate Inference 660
15.3.1 Key Ideas 661
15.3.2 Factored Belief State Methods 662
15.3.3 Particle Filtering 665
15.3.4 Deterministic Search Techniques 675
15.4 Hybrid DBNs 675
15.4.1 Continuous Models 676
15.4.2 Hybrid Models 684
15.5 Summary 688
15.6 Relevant Literature 690
15.7 Exercises 692

III Learning 695

16 Learning Graphical Models: Overview **697**
16.1 Motivation 697
16.2 Goals of Learning 698
16.2.1 Density Estimation 698
16.2.2 Specific Prediction Tasks 700
16.2.3 Knowledge Discovery 701
16.3 Learning as Optimization 702
16.3.1 Empirical Risk and Overfitting 703
16.3.2 Discriminative versus Generative Training 709
16.4 Learning Tasks 711
16.4.1 Model Constraints 712
16.4.2 Data Observability 712
16.4.3 Taxonomy of Learning Tasks 714
16.5 Relevant Literature 715

17 Parameter Estimation **717**
17.1 Maximum Likelihood Estimation 717

此为试读，需要完整PDF请访问：www.ertongbook.com