ELSEVIER
HZ BOOKS
华章教育

经 典 原 版 书 库

统计机器学习导论

[日] 杉山将（Masashi Sugiyama）著

（英文版）

INTRODUCTION TO
STATISTICAL MACHINE LEARNING

Masashi Sugiyama

机械工业出版社
China Machine Press

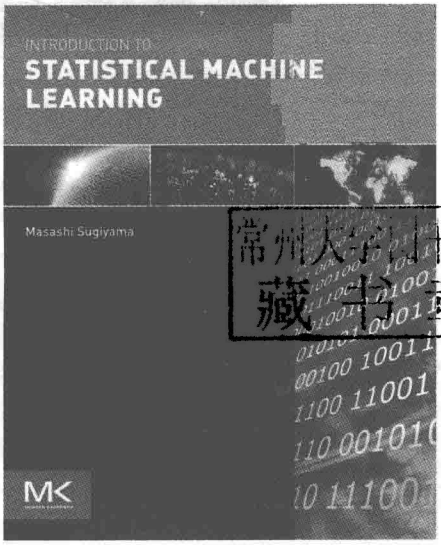
MK
MORGAN KAUFMANN

经 典 原 版 书 库

统计机器学习导论

（英文版）

Introduction to Statistical Machine Learning



常州大学图书馆
藏书章

[日] 杉山将（Masashi Sugiyama）著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

统计机器学习导论（英文版）/（日）杉山将（Masashi Sugiyama）著. —北京：机械工业出版社，2018.1
（经典原版书库）
书名原文：Introduction to Statistical Machine Learning
ISBN 978-7-111-58678-4
I. 统… II. 杉… III. 机器学习－英文 IV. TP181
中国版本图书馆 CIP 数据核字（2017）第 309485 号

本书版权登记号：图字 01-2017-7344

Elsevier (Singapore) Pte Ltd.
3 Killiney Road, #08-01 Winsland House I, Singapore 239519
Tel: (65) 6349-0200; Fax: (65) 6733-1817

Introduction to Statistical Machine Learning
Masashi Sugiyama
Copyright © 2016 Elsevier Inc. All rights of reproduction in any form reserved.
ISBN-13: 978-0-12-802121-7

This reprint of Introduction to Statistical Machine Learning by Masashi Sugiyama was undertaken by China Machine Press and is published by arrangement with Elsevier (Singapore) Pte Ltd.

Introduction to Statistical Machine Learning by Masashi Sugiyama 由机械工业出版社进行影印，并根据机械工业出版社与爱思唯尔（新加坡）私人有限公司的协议约定出版。

ISBN: 978-7-111-58678-4

Copyright © 2018 by Elsevier (Singapore) Pte Ltd.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from Elsevier (Singapore) Pte Ltd. Details on how to seek permission, further information about the Elsevier's permissions policies and arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by Elsevier (Singapore) Pte Ltd. and China Machine Press (other than as may be noted herein).

Online resources are not available with this reprint.

Notice

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors in relation to the adaptation or for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong SAR, Macau SAR and Taiwan. Unauthorized export of this edition is a violation of the contract.

本书英文影印版由 Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中华人民共和国境内（不包括香港、澳门特别行政区及台湾地区）出版与发。未经许可之出口，视为违反著作权法，将受民事和刑事法之制裁。

本书封底贴有 Elsevier 防伪标签，无标签者得销售。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：曲 熠 责任校对：殷 虹
印 刷：北京市兆成印刷有限责任公司 版 次：2018 年 1 月第 1 版第 1 次印刷
开 本：186mm×240mm 1/16 印 张：32
书 号：ISBN 978-7-111-58678-4 定 价：119.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991 88361066 投稿热线：（010）88379604
购书热线：（010）68326294 88379649 68995259 读者信箱：hzjsj@hzbook.com

版权所有 · 侵权必究
封底无防伪标均为盗版
本书法律顾问：北京大成律师事务所 韩光 / 邹晓东

出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自 1998 年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson，McGraw-Hill，Elsevier，MIT，John Wiley & Sons，Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum，Bjarne Stroustrup，Brian W. Kernighan，Dennis Ritchie，Jim Gray，Afred V. Aho，John E. Hopcroft，Jeffrey D. Ullman，Abraham Silberschatz，William Stallings，Donald E. Knuth，John L. Hennessy，Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：（010）88379604

联系地址：北京市西城区百万庄南街 1 号

邮政编码：100037

华章教育

华章科技图书出版中心

Preface

Machine learning is a subject in computer science, aimed at studying theories, algorithms, and applications of systems that learn like humans. Recent development of computers and sensors allows us to access a huge amount of data in diverse domains such as documents, audio, images, movies, e-commerce, electric power, medicine, and biology. Machine learning plays a central role in analyzing and benefiting from such *big data*.

This textbook is devoted to presenting mathematical backgrounds and practical algorithms of various machine learning techniques, targeting undergraduate and graduate students in computer science and related fields. Engineers who are applying machine learning techniques in their business and scientists who are analyzing their data can also benefit from this book.

A distinctive feature of this book is that each chapter concisely summarizes the main idea and mathematical derivation of particular machine learning techniques, followed by compact MATLAB programs. Thus, readers can study both mathematical concepts and practical values of various machine learning techniques simultaneously. All MATLAB programs are available from

"http://www.ms.k.u-tokyo.ac.jp/software/SMLbook.zip".

This book begins by giving a brief overview of the field of machine learning in Part 1. Then Part 2 introduces fundamental concepts of *probability and statistics*, which form the mathematical basis of statistical machine learning. Part 2 was written based on

Sugiyama, M.
Probability and Statistics for Machine Learning,
Kodansha, Tokyo, Japan, 2015. (in Japanese).

Part 3 and Part 4 present a variety of practical machine learning algorithms in the *generative* and *discriminative* frameworks, respectively. Then Part 5 covers various advanced topics for tackling more challenging machine learning tasks. Part 3 was written based on

Sugiyama, M.
Statistical Pattern Recognition: Pattern Recognition Based on Generative Models,
Ohmsha, Tokyo, Japan, 2009. (in Japanese),

and Part 4 and Part 5 were written based on

Sugiyama, M.
An Illustrated Guide to Machine Learning,
Kodansha, Tokyo, Japan, 2013. (in Japanese).

The author would like to thank researchers and students in his groups at the University of Tokyo and Tokyo Institute of Technology for their valuable feedback on earlier manuscripts.

Masashi Sugiyama
The University of Tokyo

试读结束：需要全本请在线购买： www.ertongbook.com

Biography

MASASHI SUGIYAMA

Masashi Sugiyama received the degrees of Bachelor of Engineering, Master of Engineering, and Doctor of Engineering in Computer Science from Tokyo Institute of Technology, Japan in 1997, 1999, and 2001, respectively. In 2001, he was appointed Assistant Professor in the same institute, and he was promoted to Associate Professor in 2003. He moved to the University of Tokyo as Professor in 2014. He received an Alexander von Humboldt Foundation Research Fellowship and researched at Fraunhofer Institute, Berlin, Germany, from 2003 to 2004. In 2006, he received a European Commission Program Erasmus Mundus Scholarship and researched at the University of Edinburgh, Edinburgh, UK. He received the Faculty Award from IBM in 2007 for his contribution to machine learning under non-stationarity, the Nagao Special Researcher Award from the Information Processing Society of Japan in 2011 and the Young Scientists' Prize from the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology Japan for his contribution to the density-ratio paradigm of machine learning. His research interests include theories and algorithms of machine learning and data mining, and a wide range of applications such as signal processing, image processing, and robot control.

Contents

Preface iv
Biography v

PART 1 INTRODUCTION

CHAPTER 1 **Statistical Machine Learning** 3
1.1 Types of Learning 3
1.2 Examples of Machine Learning Tasks 4
1.2.1 Supervised Learning 4
1.2.2 Unsupervised Learning 5
1.2.3 Further Topics 6
1.3 Structure of This Textbook 8

PART 2 STATISTICS AND PROBABILITY

CHAPTER 2 **Random Variables and Probability Distributions** 11
2.1 Mathematical Preliminaries 11
2.2 Probability 13
2.3 Random Variable and Probability Distribution 14
2.4 Properties of Probability Distributions 16
2.4.1 Expectation, Median, and Mode 16
2.4.2 Variance and Standard Deviation 18
2.4.3 Skewness, Kurtosis, and Moments 19
2.5 Transformation of Random Variables 22

CHAPTER 3 **Examples of Discrete Probability Distributions** 25
3.1 Discrete Uniform Distribution 25
3.2 Binomial Distribution 26
3.3 Hypergeometric Distribution 27
3.4 Poisson Distribution 31
3.5 Negative Binomial Distribution 33
3.6 Geometric Distribution 35

CHAPTER 4 **Examples of Continuous Probability Distributions** 37
4.1 Continuous Uniform Distribution 37
4.2 Normal Distribution 37
4.3 Gamma Distribution, Exponential Distribution, and Chi-Squared Distribution 41
4.4 Beta Distribution 44
4.5 Cauchy Distribution and Laplace Distribution 47
4.6 t-Distribution and F-Distribution 49

CHAPTER 5 **Multidimensional Probability Distributions** 51
5.1 Joint Probability Distribution 51
5.2 Conditional Probability Distribution 52
5.3 Contingency Table 53
5.4 Bayes' Theorem 53
5.5 Covariance and Correlation 55
5.6 Independence 56

CHAPTER 6 **Examples of Multidimensional Probability Distributions** 61
6.1 Multinomial Distribution 61
6.2 Multivariate Normal Distribution 62
6.3 Dirichlet Distribution 63
6.4 Wishart Distribution 70

CHAPTER 7 **Sum of Independent Random Variables** 73
7.1 Convolution 73
7.2 Reproductive Property 74
7.3 Law of Large Numbers 74
7.4 Central Limit Theorem 77

CHAPTER 8 **Probability Inequalities** 81
8.1 Union Bound 81
8.2 Inequalities for Probabilities 82
8.2.1 Markov's Inequality and Chernoff's Inequality 82
8.2.2 Cantelli's Inequality and Chebyshev's Inequality 83
8.3 Inequalities for Expectation 84
8.3.1 Jensen's Inequality 84
8.3.2 Hölder's Inequality and Schwarz's Inequality 85
8.3.3 Minkowski's Inequality 86
8.3.4 Kantorovich's Inequality 87
8.4 Inequalities for the Sum of Independent Random Variables 87
8.4.1 Chebyshev's Inequality and Chernoff's Inequality 88
8.4.2 Hoeffding's Inequality and Bernstein's Inequality 88
8.4.3 Bennett's Inequality 89

CHAPTER 9 **Statistical Estimation** 91
9.1 Fundamentals of Statistical Estimation 91
9.2 Point Estimation 92
9.2.1 Parametric Density Estimation 92
9.2.2 Nonparametric Density Estimation 93
9.2.3 Regression and Classification 93

9.2.4 Model Selection 94
9.3 Interval Estimation 95
9.3.1 Interval Estimation for Expectation of Normal Samples 95
9.3.2 Bootstrap Confidence Interval 96
9.3.3 Bayesian Credible Interval 97

CHAPTER 10 Hypothesis Testing 99
10.1 Fundamentals of Hypothesis Testing 99
10.2 Test for Expectation of Normal Samples 100
10.3 Neyman-Pearson Lemma 101
10.4 Test for Contingency Tables 102
10.5 Test for Difference in Expectations of Normal Samples 104
10.5.1 Two Samples without Correspondence 104
10.5.2 Two Samples with Correspondence 105
10.6 Nonparametric Test for Ranks 107
10.6.1 Two Samples without Correspondence 107
10.6.2 Two Samples with Correspondence 108
10.7 Monte Carlo Test 108

PART 3 GENERATIVE APPROACH TO STATISTICAL PATTERN RECOGNITION

CHAPTER 11 Pattern Recognition via Generative Model Estimation 113
11.1 Formulation of Pattern Recognition 113
11.2 Statistical Pattern Recognition 115
11.3 Criteria for Classifier Training 117
11.3.1 MAP Rule 117
11.3.2 Minimum Misclassification Rate Rule 118
11.3.3 Bayes Decision Rule 119
11.3.4 Discussion 121
11.4 Generative and Discriminative Approaches 121

CHAPTER 12 Maximum Likelihood Estimation 123
12.1 Definition 123
12.2 Gaussian Model 125
12.3 Computing the Class-Posterior Probability 127
12.4 Fisher's Linear Discriminant Analysis (FDA) 130
12.5 Hand-Written Digit Recognition 133
12.5.1 Preparation 134
12.5.2 Implementing Linear Discriminant Analysis 135
12.5.3 Multiclass Classification 136

CHAPTER 13 Properties of Maximum Likelihood Estimation 139

13.1 Consistency 139
13.2 Asymptotic Unbiasedness 140
13.3 Asymptotic Efficiency 141
13.3.1 One-Dimensional Case 141
13.3.2 Multidimensional Cases 141
13.4 Asymptotic Normality 143
13.5 Summary 145

CHAPTER 14 **Model Selection for Maximum Likelihood Estimation** **147**
14.1 Model Selection 147
14.2 KL Divergence 148
14.3 AIC 150
14.4 Cross Validation 154
14.5 Discussion 154

CHAPTER 15 **Maximum Likelihood Estimation for Gaussian Mixture Model** **157**
15.1 Gaussian Mixture Model 157
15.2 MLE 158
15.3 Gradient Ascent Algorithm 161
15.4 EM Algorithm 162

CHAPTER 16 **Nonparametric Estimation** **169**
16.1 Histogram Method 169
16.2 Problem Formulation 170
16.3 KDE 174
16.3.1 Parzen Window Method 174
16.3.2 Smoothing with Kernels 175
16.3.3 Bandwidth Selection 176
16.4 NNDE 178
16.4.1 Nearest Neighbor Distance 178
16.4.2 Nearest Neighbor Classifier 179

CHAPTER 17 **Bayesian Inference** **185**
17.1 Bayesian Predictive Distribution 185
17.1.1 Definition 185
17.1.2 Comparison with MLE 186
17.1.3 Computational Issues 188
17.2 Conjugate Prior 188
17.3 MAP Estimation 189
17.4 Bayesian Model Selection 193

CHAPTER 18 **Analytic Approximation of Marginal Likelihood** **197**
18.1 Laplace Approximation 197
18.1.1 Approximation with Gaussian Density 197

18.1.2 Illustration 199
18.1.3 Application to Marginal Likelihood Approximation 200
18.1.4 Bayesian Information Criterion (BIC) 200
18.2 Variational Approximation 202
18.2.1 Variational Bayesian EM (VBEM) Algorithm 202
18.2.2 Relation to Ordinary EM Algorithm 203

CHAPTER 19 Numerical Approximation of Predictive Distribution 205
19.1 Monte Carlo Integration 205
19.2 Importance Sampling 207
19.3 Sampling Algorithms 208
19.3.1 Inverse Transform Sampling 208
19.3.2 Rejection Sampling 212
19.3.3 Markov Chain Monte Carlo (MCMC) Method 214

CHAPTER 20 Bayesian Mixture Models 221
20.1 Gaussian Mixture Models 221
20.1.1 Bayesian Formulation 221
20.1.2 Variational Inference 223
20.1.3 Gibbs Sampling 228
20.2 Latent Dirichlet Allocation (LDA) 229
20.2.1 Topic Models 230
20.2.2 Bayesian Formulation 231
20.2.3 Gibbs Sampling 232

PART 4 DISCRIMINATIVE APPROACH TO STATISTICAL MACHINE LEARNING

CHAPTER 21 Learning Models 237
21.1 Linear-in-Parameter Model 237
21.2 Kernel Model 239
21.3 Hierarchical Model 242

CHAPTER 22 Least Squares Regression 245
22.1 Method of LS 245
22.2 Solution for Linear-in-Parameter Model 246
22.3 Properties of LS Solution 250
22.4 Learning Algorithm for Large-Scale Data 251
22.5 Learning Algorithm for Hierarchical Model 252

CHAPTER 23 Constrained LS Regression 257
23.1 Subspace-Constrained LS 257
23.2 ℓ_2-Constrained LS 259

试读结束：需要全本请在线购买： www.ertongbook.com

23.3 Model Selection 262

CHAPTER 24 **Sparse Regression** **267**
24.1 ℓ_1-Constrained LS 267
24.2 Solving ℓ_1-Constrained LS 268
24.3 Feature Selection by Sparse Learning 272
24.4 Various Extensions 272
24.4.1 Generalized ℓ_1-Constrained LS 273
24.4.2 ℓ_p-Constrained LS 273
24.4.3 $\ell_1 + \ell_2$-Constrained LS 274
24.4.4 $\ell_{1,2}$-Constrained LS 276
24.4.5 Trace Norm Constrained LS 278

CHAPTER 25 **Robust Regression** **279**
25.1 Nonrobustness of ℓ_2-Loss Minimization 279
25.2 ℓ_1-Loss Minimization 280
25.3 Huber Loss Minimization 282
25.3.1 Definition 282
25.3.2 Stochastic Gradient Algorithm 283
25.3.3 Iteratively Reweighted LS 283
25.3.4 ℓ_1-Constrained Huber Loss Minimization 286
25.4 Tukey Loss Minimization 290

CHAPTER 26 **Least Squares Classification** **295**
26.1 Classification by LS Regression 295
26.2 0/1-Loss and Margin 297
26.3 Multiclass Classification 300

CHAPTER 27 **Support Vector Classification** **303**
27.1 Maximum Margin Classification 303
27.1.1 Hard Margin Support Vector Classification 303
27.1.2 Soft Margin Support Vector Classification 305
27.2 Dual Optimization of Support Vector Classification 306
27.3 Sparseness of Dual Solution 308
27.4 Nonlinearization by Kernel Trick 311
27.5 Multiclass Extension 312
27.6 Loss Minimization View 314
27.6.1 Hinge Loss Minimization 315
27.6.2 Squared Hinge Loss Minimization 316
27.6.3 Ramp Loss Minimization 318

CHAPTER 28 **Probabilistic Classification** **321**
28.1 Logistic Regression 321
28.1.1 Logistic Model and MLE 321
28.1.2 Loss Minimization View 324

28.2 LS Probabilistic Classification 325

CHAPTER 29 Structured Classification 329
29.1 Sequence Classification 329
29.2 Probabilistic Classification for Sequences 330
29.2.1 Conditional Random Field 330
29.2.2 MLE 333
29.2.3 Recursive Computation 333
29.2.4 Prediction for New Sample 336
29.3 Deterministic Classification for Sequences 337

PART 5 FURTHER TOPICS

CHAPTER 30 Ensemble Learning 343
30.1 Decision Stump Classifier 343
30.2 Bagging 344
30.3 Boosting 346
30.3.1 Adaboost 348
30.3.2 Loss Minimization View 348
30.4 General Ensemble Learning 354

CHAPTER 31 Online Learning 355
31.1 Stochastic Gradient Descent 355
31.2 Passive-Aggressive Learning 356
31.2.1 Classification 357
31.2.2 Regression 358
31.3 Adaptive Regularization of Weight Vectors (AROW) 360
31.3.1 Uncertainty of Parameters 360
31.3.2 Classification 361
31.3.3 Regression 362

CHAPTER 32 Confidence of Prediction 365
32.1 Predictive Variance for ℓ_2-Regularized LS 365
32.2 Bootstrap Confidence Estimation 367
32.3 Applications 368
32.3.1 Time-series Prediction 368
32.3.2 Tuning Parameter Optimization 369

CHAPTER 33 Semisupervised Learning 375
33.1 Manifold Regularization 375
33.1.1 Manifold Structure Brought by Input Samples ... 375
33.1.2 Computing the Solution 377
33.2 Covariate Shift Adaptation 378
33.2.1 Importance Weighted Learning 378

33.2.2 Relative Importance Weighted Learning 382
33.2.3 Importance Weighted Cross Validation 382
33.2.4 Importance Estimation 383
33.3 Class-balance Change Adaptation 385
33.3.1 Class-balance Weighted Learning 385
33.3.2 Class-balance Estimation 386

CHAPTER 34 Multitask Learning 391
34.1 Task Similarity Regularization 391
34.1.1 Formulation 391
34.1.2 Analytic Solution 392
34.1.3 Efficient Computation for Many Tasks 393
34.2 Multidimensional Function Learning 394
34.2.1 Formulation 394
34.2.2 Efficient Analytic Solution 397
34.3 Matrix Regularization 397
34.3.1 Parameter Matrix Regularization 397
34.3.2 Proximal Gradient for Trace Norm Regularization 400

CHAPTER 35 Linear Dimensionality Reduction 405
35.1 Curse of Dimensionality 405
35.2 Unsupervised Dimensionality Reduction 407
35.2.1 PCA 407
35.2.2 Locality Preserving Projection 410
35.3 Linear Discriminant Analyses for Classification 412
35.3.1 Fisher Discriminant Analysis 413
35.3.2 Local Fisher Discriminant Analysis 414
35.3.3 Semisupervised Local Fisher Discriminant Analysis 417
35.4 Sufficient Dimensionality Reduction for Regression 419
35.4.1 Information Theoretic Formulation 419
35.4.2 Direct Derivative Estimation 422
35.5 Matrix Imputation 425

CHAPTER 36 Nonlinear Dimensionality Reduction 429
36.1 Dimensionality Reduction with Kernel Trick 429
36.1.1 Kernel PCA 429
36.1.2 Laplacian Eigenmap 433
36.2 Supervised Dimensionality Reduction with Neural Networks 435
36.3 Unsupervised Dimensionality Reduction with Autoencoder 436
36.3.1 Autoencoder 436

36.3.2 Training by Gradient Descent 437
36.3.3 Sparse Autoencoder 439
36.4 Unsupervised Dimensionality Reduction with Restricted Boltzmann Machine 440
36.4.1 Model 441
36.4.2 Training by Gradient Ascent 442
36.5 Deep Learning 446

CHAPTER 37 Clustering **447**
37.1 k-Means Clustering 447
37.2 Kernel k-Means Clustering 448
37.3 Spectral Clustering 449
37.4 Tuning Parameter Selection 452

CHAPTER 38 Outlier Detection **457**
38.1 Density Estimation and Local Outlier Factor 457
38.2 Support Vector Data Description 458
38.3 Inlier-Based Outlier Detection 464

CHAPTER 39 Change Detection **469**
39.1 Distributional Change Detection 469
39.1.1 KL Divergence 470
39.1.2 Pearson Divergence 470
39.1.3 L_2-Distance 471
39.1.4 L_1-Distance 474
39.1.5 Maximum Mean Discrepancy (MMD) 476
39.1.6 Energy Distance 477
39.1.7 Application to Change Detection in Time Series . 477
39.2 Structural Change Detection 478
39.2.1 Sparse MLE 478
39.2.2 Sparse Density Ratio Estimation 482
References 485
Index 491

PART

INTRODUCTION 1

1 STATISTICAL MACHINE LEARNING 3

PART

1

INTRODUCTION

试读结束：需要全本请在线购买： www.ertongbook.com