

模式识别

(英文版·第4版)

Pattern Recognition

Sergios Theodoridis

雅典大学

(希腊)

Konstantinos Koutroumbas

著

雅典国家天文台
空间应用研究院



机械工业出版社
China Machine Press

经典原版书库

模式识别

(英文版·第4版)

Pattern Recognition

江苏工业学院图书馆
藏书章



机械工业出版社
China Machine Press

Sergios Theodoridis and Konstantinos Koutroumbas; Pattern Recognition, Fourth Edition (ISBN 978-1-59749-272-0).

Original English language edition copyright © 2009 by Elsevier Inc. All rights reserved.

Authorized English language reprint edition published by the Proprietor.

ISBN:978-981-272-337-6

Copyright © 2009 by Elsevier (Singapore) Pte Ltd.

Printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in China only, excluding Hong Kong SAR and Taiwan.

Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书英文影印版由Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中国大陆境内独家发行。本版仅限在中国境内（不包括中国香港特别行政区及中国台湾地区）出版及标价销售。未经许可之出口，视为违反著作权法，将受法律之制裁。

版权所有，侵权必究。

本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2009-2372

图书在版编目（CIP）数据

模式识别（英文版·第4版）/（希）西奥多里德斯（Theodoridis, S.）等著.
—北京：机械工业出版社，2009.8

（经典原版书库）

书名原书：Pattern Recognition, Fourth Edition

ISBN 978-7-111-26896-3

I. 模… II. 西… III. 模式识别—英文 IV. TP391.4

中国版本图书馆CIP数据核字（2009）第061214号

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：李东震

北京京师印务有限公司印刷

2009年8月第1版第1次印刷

150mm×214mm · 30.625印张

标准书号：ISBN 978-7-111-26896-3

定价：89.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换
本社购书热线：（010）68326294

出版者的话

文艺复兴以降，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；也正是这样的传统，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，不仅肇划了研究的范畴，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分社较早意识到“出版要为教育服务”。自1998年开始，华章分社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, Brain 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襄助，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，有的还专程为其

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。华章分社欢迎老师和读者对我们的工作提出建议或给予指正,我们的联系方式如下:

华章网站: www.hzbook.com

电子邮件: hzsj@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



Preface

This book is the outgrowth of our teaching advanced undergraduate and graduate courses over the past 20 years. These courses have been taught to different audiences, including students in electrical and electronics engineering, computer engineering, computer science, and informatics, as well as to an interdisciplinary audience of a graduate course on automation. This experience led us to make the book as self-contained as possible and to address students with different backgrounds. As prerequisite knowledge, the reader requires only basic calculus, elementary linear algebra, and some probability theory basics. A number of mathematical tools, such as probability and statistics as well as constrained optimization, needed by various chapters, are treated in four Appendices. The book is designed to serve as a text for advanced undergraduate and graduate students, and it can be used for either a one- or a two-semester course. Furthermore, it is intended to be used as a self-study and reference book for research and for the practicing scientist/engineer. This latter audience was also our second incentive for writing this book, due to the involvement of our group in a number of projects related to pattern recognition.

SCOPE AND APPROACH

The goal of the book is to present in a unified way the most widely used techniques and methodologies for pattern recognition tasks. Pattern recognition is in the center of a number of application areas, including image analysis, speech and audio recognition, biometrics, bioinformatics, data mining, and information retrieval. Despite their differences, these areas share, to a large extent, a corpus of techniques that can be used in extracting, from the available data, information related to data categories, important "hidden" patterns, and trends. The emphasis in this book is on the most generic of the methods that are currently available. Having acquired the basic knowledge and understanding, the reader can subsequently move on to more specialized application-dependent techniques, which have been developed and reported in a vast number of research papers.

Each chapter of the book starts with the basics and moves, progressively, to more advanced topics and reviews up-to-date techniques. We have made an effort to keep a balance between mathematical and descriptive presentation. This is not always an easy task. However, we strongly believe that in a topic such as pattern recognition, trying to bypass mathematics deprives the reader of understanding the essentials behind the methods and also the potential of developing new techniques, which fit the needs of the problem at hand that he or she has to tackle. In pattern recognition, the final adoption of an appropriate technique and algorithm is very much a problem-dependent task. Moreover, according to our experience, teaching pattern recognition is also a good "excuse" for the students to refresh and solidify

some of the mathematical basics they have been taught in earlier years. "Repetitio est mater studiosum."

NEW TO THIS EDITION

The new features of the fourth edition include the following.

- MATLAB codes and computer experiments are given at the end of most chapters.
- More examples and a number of new figures have been included to enhance the readability and pedagogic aspects of the book.
- New sections on some important topics of high current interest have been added, including:
 - Nonlinear dimensionality reduction
 - Nonnegative matrix factorization
 - Relevance feedback
 - Robust regression
 - Semi-supervised learning
 - Spectral clustering
 - Clustering combination techniques

Also, a number of sections have been rewritten in the context of more recent applications in mind.

SUPPLEMENTS TO THE TEXT

Demonstrations based on MATLAB are available for download from the book Web site, www.elsevierdirect.com/9781597492720. Also available are electronic figures from the text and (for instructors only) a solutions manual for the end-of-chapter problems and exercises. The interested reader can download detailed proofs, which in the book necessarily, are sometimes, slightly condensed. PowerPoint presentations are also available covering all chapters of the book.

Our intention is to update the site regularly with more and/or improved versions of the MATLAB demonstrations. Suggestions are always welcome. Also at this Web site a page will be available for typos, which are unavoidable, despite frequent careful reading. The authors would appreciate readers notifying them about any typos found.

ACKNOWLEDGMENTS

This book would have not been written without the constant support and help from a number of colleagues and students throughout the years. We are especially indebted to Kostas Berberidis, Velissaris Gezerlis, Xaris Georgion, Kristina Georgoulakis, Leyteris Kofidis, Thanassis Liavas, Michalis Mavroforakis, Aggelos Pikrakis, Thanassis Rontogiannis, Margaritis Sdralis, Kostas Slavakis, and Theodoros Yiannakoponlos. The constant support provided by Yannis Kopsinis and Kostas Thernelis from the early stages up to the final stage, with those long nights, has been invaluable. The book improved a great deal after the careful reading and the serious comments and suggestions of Alexandros Bölnn. Dionissis Cavouras, Vassilis Digalakis, Vassilis Drakopoulos, Nikos Galatsanos, George Glentis, Spiros Hatzispyros, Evangelos Karkaletsis, Elias Koutsoupas, Aristides Likas, Gerassimos Mileounis, George Monstakides, George Paliouras, Stavros Perantonis, Takis Stamatooponlos, Nikos Vassilas, Manolis Zervakis, and Vassilis Zissimopoulos.

The book has greatly gained and improved thanks to the comments of a number of people who provided feedback on the revision plan and/or comments on revised chapters:

Tulay Adali, University of Maryland; Mehnet Celenk, Ohio University; Rama Chelappa, University of Maryland; Mark Clements, Georgia Institute of Technology; Robert Duin, Delft University of Technology; Miguel Figneroa, Villanueva University of Puerto Rico; Dimitris Gunopoulos, University of Athens; Mathias Kolsch, Naval Postgraduate School; Adam Krzyzak, Concordia University; Baoxiu Li, Arizona State University; David Miller, Pennsylvania State University; Bernhard Schölkopf, Max Planck Institute; Hari Sundaram, Arizona State University; Harry Wechsler, George Mason University; and Alexander Zien, Max Planck Institute.

We are greatly indebted to these colleagues for their time and their constructive criticisms. Our collaboration and friendship with Nikos Kalouptsidis have been a source of constant inspiration for all these years. We are both deeply indebted to him.

Last but not least, K. Koutroumbas would like to thank Sophia, Dimitris-Marios, and Valentini-Theodora for their tolerance and support and S.Theodoridis would like to thank Despina, Eva, and Eleni, his joyful and supportive "barem."

Contents

Preface	v
CHAPTER 1 Introduction	1
1.1 Is Pattern Recognition Important?	1
1.2 Features, Feature Vectors, and Classifiers	4
1.3 Supervised, Unsupervised, and Semi-Supervised Learning	7
1.4 MATLAB Programs	9
1.5 Outline of The Book	10
CHAPTER 2 Classifiers Based on Bayes Decision Theory	13
2.1 Introduction	13
2.2 Bayes Decision Theory	13
2.3 Discriminant Functions and Decision Surfaces	19
2.4 Bayesian Classification for Normal Distributions	20
2.4.1 The Gaussian Probability Density Function	20
2.4.2 The Bayesian Classifier for Normally Distributed Classes	24
2.5 Estimation of Unknown Probability Density Functions	34
2.5.1 Maximum Likelihood Parameter Estimation	34
2.5.2 Maximum <i>a Posteriori</i> Probability Estimation	38
2.5.3 Bayesian Inference	39
2.5.4 Maximum Entropy Estimation	43
2.5.5 Mixture Models	44
2.5.6 Nonparametric Estimation	49
2.5.7 The Naive-Bayes Classifier	59
2.6 The Nearest Neighbor Rule	61
2.7 Bayesian Networks	64
2.8 Problems	71
References	86
CHAPTER 3 Linear Classifiers	91
3.1 Introduction	91
3.2 Linear Discriminant Functions and Decision Hyperplanes	91
3.3 The Perceptron Algorithm	93
3.4 Least Squares Methods	103
3.4.1 Mean Square Error Estimation	103

3.4.2	Stochastic Approximation and the LMS Algorithm ..	105
3.4.3	Sum of Error Squares Estimation	108
3.5	Mean Square Estimation Revisited	110
3.5.1	Mean Square Error Regression	110
3.5.2	MSE Estimates Posterior Class Probabilities	112
3.5.3	The Bias-Variance Dilemma	114
3.6	Logistic Discrimination	117
3.7	Support Vector Machines	119
3.7.1	Separable Classes	119
3.7.2	Nonseparable Classes	124
3.7.3	The Multiclass Case	127
3.7.4	ν -SVM	133
3.7.5	Support Vector Machines: A Geometric Viewpoint	136
3.7.6	Reduced Convex Hulls	138
3.8	Problems	142
	References	147
CHAPTER 4	Nonlinear Classifiers	151
4.1	Introduction	151
4.2	The XOR Problem	151
4.3	The Two-Layer Perceptron	153
4.3.1	Classification Capabilities of the Two-Layer Perceptron	156
4.4	Three-Layer Perceptrons	158
4.5	Algorithms Based on Exact Classification of the Training Set	160
4.6	The Backpropagation Algorithm	162
4.7	Variations on the Backpropagation Theme	169
4.8	The Cost Function Choice	172
4.9	Choice of the Network Size	176
4.10	A Simulation Example	181
4.11	Networks with Weight Sharing	183
4.12	Generalized Linear Classifiers	185
4.13	Capacity of the I -Dimensional Space in Linear Dichotomies	187
4.14	Polynomial Classifiers	189
4.15	Radial Basis Function Networks	190
4.16	Universal Approximators	194
4.17	Probabilistic Neural Networks	196
4.18	Support Vector Machines: The Nonlinear Case	198

4.19	Beyond the SVM Paradigm	203
4.19.1	Expansion in Kernel Functions and Model Sparsification	205
4.19.2	Robust Statistics Regression	211
4.20	Decision Trees	215
4.20.1	Set of Questions	218
4.20.2	Splitting Criterion	218
4.20.3	Stop-Splitting Rule	219
4.20.4	Class Assignment Rule	219
4.21	Combining Classifiers	222
4.21.1	Geometric Average Rule	223
4.21.2	Arithmetic Average Rule	224
4.21.3	Majority Voting Rule	225
4.21.4	A Bayesian Viewpoint	227
4.22	The Boosting Approach to Combine Classifiers	230
4.23	The Class Imbalance Problem	237
4.24	Discussion	239
4.25	Problems	240
	References	249
CHAPTER 5	Feature Selection	261
5.1	Introduction	261
5.2	Preprocessing	262
5.2.1	Outlier Removal	262
5.2.2	Data Normalization	263
5.2.3	Missing Data	263
5.3	The Peaking Phenomenon	265
5.4	Feature Selection Based on Statistical Hypothesis Testing	268
5.4.1	Hypothesis Testing Basics	268
5.4.2	Application of the t -Test in Feature Selection	273
5.5	The Receiver Operating Characteristics (ROC) Curve	275
5.6	Class Separability Measures	276
5.6.1	Divergence	276
5.6.2	Chernoff Bound and Bhattacharyya Distance	278
5.6.3	Scatter Matrices	280
5.7	Feature Subset Selection	283
5.7.1	Scalar Feature Selection	283
5.7.2	Feature Vector Selection	284
5.8	Optimal Feature Generation	288
5.9	Neural Networks and Feature Generation/Selection	298
5.10	A Hint On Generalization Theory	299

5.11	The Bayesian Information Criterion	309
5.12	Problems	311
	References	318
CHAPTER 6	Feature Generation I: Data Transformation and Dimensionality Reduction	323
6.1	Introduction	323
6.2	Basis Vectors and Images	324
6.3	The Karhunen-Loève Transform	326
6.4	The Singular Value Decomposition	335
6.5	Independent Component Analysis	342
6.5.1	ICA Based on Second- and Fourth-Order Cumulants	344
6.5.2	ICA Based on Mutual Information	345
6.5.3	An ICA Simulation Example	348
6.6	Nonnegative Matrix Factorization	349
6.7	Nonlinear Dimensionality Reduction	350
6.7.1	Kernel PCA	351
6.7.2	Graph-Based Methods	353
6.8	The Discrete Fourier Transform (DFT)	363
6.8.1	One-Dimensional DFT	364
6.8.2	Two-Dimensional DFT	366
6.9	The Discrete Cosine and Sine Transforms	366
6.10	The Hadamard Transform	368
6.11	The Haar Transform	369
6.12	The Haar Expansion Revisited	371
6.13	Discrete Time Wavelet Transform (DTWT)	375
6.14	The Multiresolution Interpretation	384
6.15	Wavelet Packets	387
6.16	A Look at Two-Dimensional Generalizations	388
6.17	Applications	390
6.18	Problems	396
	References	402
CHAPTER 7	Feature Generation II	411
7.1	Introduction	411
7.2	Regional Features	412
7.2.1	Features for Texture Characterization	412
7.2.2	Local Linear Transforms for Texture Feature Extraction	421
7.2.3	Moments	423
7.2.4	Parametric Models	427

7.3	Features for Shape and Size Characterization	435
7.3.1	Fourier Features	436
7.3.2	Chain Codes	439
7.3.3	Moment-Based Features	441
7.3.4	Geometric Features	442
7.4	A Glimpse at Fractals	444
7.4.1	Self-Similarity and Fractal Dimension	444
7.4.2	Fractional Brownian Motion	446
7.5	Typical Features for Speech and Audio Classification	451
7.5.1	Short Time Processing of Signals	452
7.5.2	Cepstrum	455
7.5.3	The Mel-Cepstrum	457
7.5.4	Spectral Features	460
7.5.5	Time Domain Features	462
7.5.6	An Example	463
7.6	Problems	466
	References	473
CHAPTER 8	Template Matching	481
8.1	Introduction	481
8.2	Measures Based on Optimal Path Searching Techniques	482
8.2.1	Bellman's Optimality Principle and Dynamic Programming	484
8.2.2	The Edit Distance	487
8.2.3	Dynamic Time Warping in Speech Recognition	491
8.3	Measures Based on Correlations	498
8.4	Deformable Template Models	504
8.5	Content-Based Information Retrieval:	
	Relevance Feedback	508
8.6	Problems	513
	References	517
CHAPTER 9	Context-Dependent Classification	521
9.1	Introduction	521
9.2	The Bayes Classifier	521
9.3	Markov Chain Models	522
9.4	The Viterbi Algorithm	523
9.5	Channel Equalization	527
9.6	Hidden Markov Models	532
9.7	HMM with State Duration Modeling	545
9.8	Training Markov Models via Neural Networks	552
9.9	A discussion of Markov Random Fields	554

9.10	Problems	556
	References	560
CHAPTER 10	Supervised Learning: The Epilogue	567
10.1	Introduction	567
10.2	Error-Counting Approach	568
10.3	Exploiting the Finite Size of the Data Set	569
10.4	A Case Study from Medical Imaging	573
10.5	Semi-Supervised Learning	577
10.5.1	Generative Models	579
10.5.2	Graph-Based Methods	582
10.5.3	Transductive Support Vector Machines	586
10.6	Problems	590
	References	591
CHAPTER 11	Clustering: Basic Concepts	595
11.1	Introduction	595
11.1.1	Applications of Cluster Analysis	598
11.1.2	Types of Features	599
11.1.3	Definitions of Clustering	600
11.2	Proximity Measures	602
11.2.1	Definitions	602
11.2.2	Proximity Measures between Two Points	604
11.2.3	Proximity Functions between a Point and a Set	616
11.2.4	Proximity Functions between Two Sets	620
11.3	Problems	622
	References	624
CHAPTER 12	Clustering Algorithms I: Sequential Algorithms	627
12.1	Introduction	627
12.1.1	Number of Possible Clusterings	627
12.2	Categories of Clustering Algorithms	629
12.3	Sequential Clustering Algorithms	633
12.3.1	Estimation of the Number of Clusters	635
12.4	A Modification of BSAS	637
12.5	A Two-Threshold Sequential Scheme	638
12.6	Refinement Stages	641
12.7	Neural Network Implementation	643
12.7.1	Description of the Architecture	643
12.7.2	Implementation of the BSAS Algorithm	644

12.8	Problems	646
	References	650
CHAPTER 13	Clustering Algorithms II: Hierarchical Algorithms	653
13.1	Introduction	653
13.2	Agglomerative Algorithms	654
13.2.1	Definition of Some Useful Quantities	655
13.2.2	Agglomerative Algorithms Based on Matrix Theory	658
13.2.3	Monotonicity and Crossover	664
13.2.4	Implementational Issues	667
13.2.5	Agglomerative Algorithms Based on Graph Theory	667
13.2.6	Ties in the Proximity Matrix	676
13.3	The Cophenetic Matrix	679
13.4	Divisive Algorithms	680
13.5	Hierarchical Algorithms for Large Data Sets	682
13.6	Choice of the Best Number of Clusters	690
13.7	Problems	693
	References	697
CHAPTER 14	Clustering Algorithms III: Schemes Based on Function Optimization	701
14.1	Introduction	701
14.2	Mixture Decomposition Schemes	703
14.2.1	Compact and Hyperellipsoidal Clusters	705
14.2.2	A Geometrical Interpretation	709
14.3	Fuzzy Clustering Algorithms	712
14.3.1	Point Representatives	716
14.3.2	Quadric Surfaces as Representatives	718
14.3.3	Hyperplane Representatives	728
14.3.4	Combining Quadric and Hyperplane Representatives	731
14.3.5	A Geometrical Interpretation	732
14.3.6	Convergence Aspects of the Fuzzy Clustering Algorithms	732
14.3.7	Alternating Cluster Estimation	733
14.4	Possibilistic Clustering	733
14.4.1	The Mode-Seeking Property	737
14.4.2	An Alternative Possibilistic Scheme	739
14.5	Hard Clustering Algorithms	739
14.5.1	The Isodata or k-Means or c-Means Algorithm	741
14.5.2	k-Medoids Algorithms	745
14.6	Vector Quantization	749

14.7	Problems	752
	References	758
CHAPTER 15	Clustering Algorithms IV	765
15.1	Introduction	765
15.2	Clustering Algorithms Based on Graph Theory	765
15.2.1	Minimum Spanning Tree Algorithms	766
15.2.2	Algorithms Based on Regions of Influence	768
15.2.3	Algorithms Based on Directed Trees	770
15.2.4	Spectral Clustering	772
15.3	Competitive Learning Algorithms	780
15.3.1	Basic Competitive Learning Algorithm	782
15.3.2	Leaky Learning Algorithm	783
15.3.3	Conscientious Competitive Learning Algorithms	784
15.3.4	Competitive Learning-Like Algorithms Associated with Cost Functions	785
15.3.5	Self-Organizing Maps	786
15.3.6	Supervised Learning Vector Quantization	788
15.4	Binary Morphology Clustering Algorithms (BMCAs)	789
15.4.1	Discretization	790
15.4.2	Morphological Operations	791
15.4.3	Determination of the Clusters in a Discrete Binary Set	794
15.4.4	Assignment of Feature Vectors to Clusters	795
15.4.5	The Algorithmic Scheme	796
15.5	Boundary Detection Algorithms	798
15.6	Valley-Seeking Clustering Algorithms	801
15.7	Clustering via Cost Optimization (Revisited)	803
15.7.1	Branch and Bound Clustering Algorithms	803
15.7.2	Simulated Annealing	807
15.7.3	Deterministic Annealing	808
15.7.4	Clustering Using Genetic Algorithms	810
15.8	Kernel Clustering Methods	811
15.9	Density-Based Algorithms for Large Data Sets	815
15.9.1	The DBSCAN Algorithm	815
15.9.2	The DBCLASD Algorithm	818
15.9.3	The DENCLUE Algorithm	819
15.10	Clustering Algorithms for High-Dimensional Data Sets	821
15.10.1	Dimensionality Reduction Clustering Approach	822
15.10.2	Subspace Clustering Approach	824
15.11	Other Clustering Algorithms	837
15.12	Combination of Clusterings	839

15.13 Problems	846
References	852
CHAPTER 16 Cluster Validity	863
16.1 Introduction	863
16.2 Hypothesis Testing Revisited	864
16.3 Hypothesis Testing in Cluster Validity	866
16.3.1 External Criteria	868
16.3.2 Internal Criteria	873
16.4 Relative Criteria	877
16.4.1 Hard Clustering	880
16.4.2 Fuzzy Clustering	887
16.5 Validity of Individual Clusters	893
16.5.1 External Criteria	894
16.5.2 Internal Criteria	894
16.6 Clustering Tendency	896
16.6.1 Tests for Spatial Randomness	900
16.7 Problems	905
References	909
APPENDIX A Hints from Probability and Statistics	915
APPENDIX B Linear Algebra Basics	927
APPENDIX C Cost Function Optimization	930
APPENDIX D Basic Definitions from Linear Systems Theory	946
Index	949