

# 模式识别

(英文版·第3版)

## PATTERN RECOGNITION

T H I R D E D I T I O N

SERGIOS  
THEODORIDIS  
KONSTANTINOS  
KOUTROUMBAS

(希腊) Sergios Theodoridis  
Konstantinos Koutroumbas 著



机械工业出版社  
China Machine Press

经 典 原 版 书 库

# 模式识别

(英文版 · 第3版)

Pattern Recognition  
(Third Edition)

江苏工业学院图书馆  
藏书章

(希腊) Sergios Theodoridis  
Konstantinos Koutroumbas 著



机械工业出版社  
China Machine Press

Sergios Theodoridis and Konstantinos Koutroumbas: Pattern Recognition, Third Edition (ISBN:0-12-369531-7 ISBN-13: 978-0-12-369531-4).

Original English language edition copyright © 2006 by Elsevier (USA). All rights reserved.

Authorized English language reprint edition published by the Proprietor.

ISBN: 981-259-706-9 ISBN-13: 978-981-259-706-9

Copyright © 2006 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（不包括香港特别行政区及台湾地区）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：01-2006-3123**

**图书在版编目（CIP）数据**

模式识别（英文版·第3版）/（希腊）西奥多里迪斯（Theodoridis, S.）等著. —北京：机械工业出版社，2006.9

（经典原版书库）

书名原文：Pattern Recognition, Third Edition

ISBN 7-111-19767-4

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（2006）第096646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：迟振春

北京京北制版印刷厂印刷·新华书店北京发行所发行

2006年9月第1版第1次印刷

170mm × 242mm · 53.5印张

定价：79.00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，还揭橥了学术的源变，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、从业人员较少的现状下，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图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall, Addison-Wesley, McGraw-Hill, Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum, Stroustrup, Kernighan, Jim Gray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、研究及度藏。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，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步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步入一个新的阶段。为此，华章公司将加大引进教材的力度，在“华章教育”的总规划之下出版三个系列的计算机教材：除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”；同时，引进全美通行的教学辅导书“Schaum's Outlines”系列组成“全美经典学习指导系列”。为了保证

这三套丛书的权威性，同时也为了更好地为学校和老师服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这三套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业的教学度身订造的。其中许多教材均已为M. I. T., Stanford, U.C. Berkeley, C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

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，但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。教材的出版只是我们的后续服务的起点。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，我们的联系方式如下：

电子邮件: [hzjsj@hzbook.com](mailto:hzjsj@hzbook.com)

联系电话: (010) 68995264

联系地址: 北京市西城区百万庄南街1号

邮政编码: 100037

# 专家指导委员会

(按姓氏笔画顺序)

尤晋元  
石教英  
张立昂  
邵维忠  
周克定  
郑国梁  
高传善  
裘宗燕

王 珊  
吕 建  
李伟琴  
陆丽娜  
周傲英  
施伯乐  
梅 宏  
戴 葵

冯博琴  
孙玉芳  
李师贤  
陆鑫达  
孟小峰  
钟玉琢  
程 旭

史忠植  
吴世忠  
李建中  
陈向群  
岳丽华  
唐世渭  
程时端

史美林  
吴时霖  
杨冬青  
周伯生  
范 明  
袁崇义  
谢希仁

# 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.

The philosophy of the book is to present various pattern recognition tasks in a unified way, including image analysis, speech processing, and communication applications. Despite their differences, these areas do share common features and their study can only benefit from a unified approach. Each chapter of the book starts with the basics and moves progressively to more advanced topics and reviews up-to-date techniques. A number of problems and computer exercises are given at the end of each chapter and a solutions manual is available from the publisher. Furthermore, a number of demonstrations based on MATLAB are available via the web at the book's site, <http://www.di.uoa.gr/~stpatrec>.

Our intention is to update the site regularly with more and/or improved versions of these 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.

This book would have not be written without the constant support and help from a number of colleagues and students throughout the years. We are especially indebted to Prof. K. Berberidis, Dr. E. Kofidis, Prof. A. Liavas, Dr. A. Rontogiannis, Dr. A. Pikrakis, Dr. Gezerlis and Dr. K. Georgoulakis. The constant support provided by Dr. I. Kopsinis 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 Prof. G. Moustakides, Prof. V. Digalakis, Prof. T. Adali, Prof. M. Zervakis, Prof. D. Cavouras, Prof. A. Böhm, Prof. G. Glentis, Prof. E. Koutsoupias, Prof. V. Zissimopoulos, Prof. A. Likas, Dr. S. Perantonis, Dr. A. Vassiliou, Dr. N. Vassilas, Dr. V. Drakopoulos, Dr. S. Hatzispyros, Prof. T. Stamatopoulos, Dr. G. Paliouras, Dr. V. Karkaletsis, Dr. Mavroforakis, Dr. S. Dralis. We are greatly indebted to these colleagues for their time and their constructive criticisms. Our collaboration and friendship with Prof. N. 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-Marius, and Valentini-Theodora for their tolerance and support and S. Theodoridis would like to thank Despina, Eva, and Eleni, his joyful and supportive "harem."*

# CONTENTS

Preface	vii
CHAPTER 1 INTRODUCTION	1
1.1 Is Pattern Recognition Important?	1
1.2 Features, Feature Vectors, and Classifiers	3
1.3 Supervised Versus Unsupervised Pattern Recognition	6
1.4 Outline of the Book	8
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
2.5 Estimation of Unknown Probability Density Functions	28
2.5.1 Maximum Likelihood Parameter Estimation	28
2.5.2 Maximum <i>a Posteriori</i> Probability Estimation	32
2.5.3 Bayesian Inference	33
2.5.4 Maximum Entropy Estimation	35
2.5.5 Mixture Models	36
2.5.6 Nonparametric Estimation	42
2.5.7 The Naive-Bayes Classifier	47
2.6 The Nearest Neighbor Rule	48
2.7 Bayesian Networks	50
CHAPTER 3 LINEAR CLASSIFIERS	69
3.1 Introduction	69
3.2 Linear Discriminant Functions and Decision Hyperplanes	69

3.3	The Perceptron Algorithm	71
3.4	Least Squares Methods	79
3.4.1	Mean Square Error Estimation	79
3.4.2	Stochastic Approximation and the LMS Algorithm	82
3.4.3	Sum of Error Squares Estimation	84
3.5	Mean Square Estimation Revisited	86
3.5.1	Mean Square Error Regression	86
3.5.2	MSE Estimates Posterior Class Probabilities	87
3.5.3	The Bias–Variance Dilemma	90
3.6	Logistic Discrimination	91
3.7	Support Vector Machines	93
3.7.1	Separable Classes	93
3.7.2	Nonseparable Classes	98
3.7.3	$\nu$ -SVM	106
3.7.4	Support Vector Machines: A Geometric Viewpoint	110
3.7.5	Reduced Convex Hulls	112
CHAPTER 4	NONLINEAR CLASSIFIERS	121
4.1	Introduction	121
4.2	The XOR Problem	121
4.3	The Two-Layer Perceptron	122
4.3.1	Classification Capabilities of the Two-Layer Perceptron	126
4.4	Three-Layer Perceptrons	129
4.5	Algorithms Based on Exact Classification of the Training Set	130
4.6	The Backpropagation Algorithm	132
4.7	Variations on the Backpropagation Theme	140
4.8	The Cost Function Choice	143
4.9	Choice of the Network Size	147
4.10	A Simulation Example	153
4.11	Networks With Weight Sharing	155
4.12	Generalized Linear Classifiers	156
4.13	Capacity of the $l$ -Dimensional Space in Linear Dichotomies	158
4.14	Polynomial Classifiers	161
4.15	Radial Basis Function Networks	162
4.16	Universal Approximators	167
4.17	Support Vector Machines: The Nonlinear Case	169

4.18	Decision Trees	174
4.18.1	Set of Questions	176
4.18.2	Splitting Criterion	177
4.18.3	Stop-Splitting Rule	178
4.18.4	Class Assignment Rule	178
4.19	Combining Classifiers	181
4.19.1	Geometric Average Rule	182
4.19.2	Arithmetic Average Rule	182
4.19.3	Majority Voting Rule	183
4.19.4	A Bayesian Viewpoint	185
4.20	The Boosting Approach to Combine Classifiers	188
4.21	Discussion	196
CHAPTER 5	FEATURE SELECTION	213
5.1	Introduction	213
5.2	Preprocessing	214
5.2.1	Outlier Removal	214
5.2.2	Data Normalization	215
5.2.3	Missing Data	215
5.3	Feature Selection Based on Statistical Hypothesis Testing	216
5.3.1	Hypothesis Testing Basics	216
5.3.2	Application of the $t$ -Test in Feature Selection	221
5.4	The Receiver Operating Characteristics (ROC) Curve	223
5.5	Class Separability Measures	224
5.5.1	Divergence	225
5.5.2	Chernoff Bound and Bhattacharyya Distance	227
5.5.3	Scatter Matrices	228
5.6	Feature Subset Selection	231
5.6.1	Scalar Feature Selection	232
5.6.2	Feature Vector Selection	233
5.7	Optimal Feature Generation	237
5.8	Neural Networks and Feature Generation/Selection	242
5.9	A Hint on Generalization Theory	243
5.10	The Bayesian Information Criterion	253
CHAPTER 6	FEATURE GENERATION I: LINEAR TRANSFORMS	263
6.1	Introduction	263
6.2	Basis Vectors and Images	264

6.3	The Karhunen–Loève Transform	266
6.4	The Singular Value Decomposition	273
6.5	Independent Component Analysis	276
6.5.1	ICA Based on Second- and Fourth-Order Cumulants	278
6.5.2	ICA Based on Mutual Information	280
6.5.3	An ICA Simulation Example	283
6.6	The Discrete Fourier Transform (DFT)	285
6.6.1	One-Dimensional DFT	285
6.6.2	Two-Dimensional DFT	287
6.7	The Discrete Cosine and Sine Transforms	288
6.8	The Hadamard Transform	290
6.9	The Haar Transform	291
6.10	The Haar Expansion Revisited	292
6.11	Discrete Time Wavelet Transform (DTWT)	297
6.12	The Multiresolution Interpretation	307
6.13	Wavelet Packets	309
6.14	A Look at Two-Dimensional Generalizations	311
6.15	Applications	313
CHAPTER 7	FEATURE GENERATION II	327
7.1	Introduction	327
7.2	Regional Features	328
7.2.1	Features for Texture Characterization	328
7.2.2	Local Linear Transforms for Texture Feature Extraction	337
7.2.3	Moments	340
7.2.4	Parametric Models	344
7.3	Features for Shape and Size Characterization	353
7.3.1	Fourier Features	354
7.3.2	Chain Codes	357
7.3.3	Moment-Based Features	359
7.3.4	Geometric Features	360
7.4	A Glimpse at Fractals	362
7.4.1	Self-Similarity and Fractal Dimension	362
7.4.2	Fractional Brownian Motion	365
7.5	Typical Features for Speech and Audio Classification	370
7.5.1	Short Time Processing of Signals	371
7.5.2	Cepstrum	375
7.5.3	The Mel-Cepstrum	377
7.5.4	Special Features	381

7.5.5	Time Domain Features	383
7.5.6	An Example	384
CHAPTER 8	TEMPLATE MATCHING	397
8.1	Introduction	397
8.2	Measures Based on Optimal Path Searching Techniques	398
8.2.1	Bellman's Optimality Principle and Dynamic Programming	400
8.2.2	The Edit Distance	401
8.2.3	Dynamic Time Warping in Speech Recognition	406
8.3	Measures Based on Correlations	413
8.4	Deformable Template Models	419
CHAPTER 9	CONTEXT-DEPENDENT CLASSIFICATION	427
9.1	Introduction	427
9.2	The Bayes Classifier	427
9.3	Markov Chain Models	428
9.4	The Viterbi Algorithm	429
9.5	Channel Equalization	432
9.6	Hidden Markov Models	437
9.7	HMM with State Duration Modeling	452
9.8	Training Markov Models via Neural Networks	458
9.9	A Discussion of Markov Random Fields	460
CHAPTER 10	SYSTEM EVALUATION	471
10.1	Introduction	471
10.2	Error Counting Approach	471
10.3	Exploiting the Finite Size of the Data Set	473
10.4	A Case Study From Medical Imaging	476
CHAPTER 11	CLUSTERING: BASIC CONCEPTS	483
11.1	Introduction	483
11.1.1	Applications of Cluster Analysis	486
11.1.2	Types of Features	487
11.1.3	Definitions of Clustering	488
11.2	Proximity Measures	490
11.2.1	Definitions	490
11.2.2	Proximity Measures between Two Points	493

11.2.3	Proximity Functions between a Point and a Set	505
11.2.4	Proximity Functions between Two Sets	510
CHAPTER 12	CLUSTERING ALGORITHMS I: SEQUENTIAL ALGORITHMS	517
12.1	Introduction	517
12.1.1	Number of Possible Clusterings	517
12.2	Categories of Clustering Algorithms	519
12.3	Sequential Clustering Algorithms	523
12.3.1	Estimation of the Number of Clusters	525
12.4	A Modification of BSAS	527
12.5	A Two-Threshold Sequential Scheme	529
12.6	Refinement Stages	531
12.7	Neural Network Implementation	533
12.7.1	Description of the Architecture	533
12.7.2	Implementation of the BSAS Algorithm	535
CHAPTER 13	CLUSTERING ALGORITHMS II: HIERARCHICAL ALGORITHMS	541
13.1	Introduction	541
13.2	Agglomerative Algorithms	542
13.2.1	Definition of Some Useful Quantities	543
13.2.2	Agglomerative Algorithms Based on Matrix Theory	545
13.2.3	Monotonicity and Crossover	553
13.2.4	Implementational Issues	556
13.2.5	Agglomerative Algorithms Based on Graph Theory	556
13.2.6	Ties in the Proximity Matrix	565
13.3	The Cophenetic Matrix	568
13.4	Divisive Algorithms	570
13.5	Hierarchical Algorithms for Large Data Sets	572
13.6	Choice of the Best Number of Clusters	580
CHAPTER 14	CLUSTERING ALGORITHMS III: SCHEMES BASED ON FUNCTION OPTIMIZATION	589
14.1	Introduction	589
14.2	Mixture Decomposition Schemes	591
14.2.1	Compact and Hyperellipsoidal Clusters	593
14.2.2	A Geometrical Interpretation	597

14.3	Fuzzy Clustering Algorithms	600
14.3.1	Point Representatives	605
14.3.2	Quadric Surfaces as Representatives	607
14.3.3	Hyperplane Representatives	617
14.3.4	Combining Quadric and Hyperplane Representatives	619
14.3.5	A Geometrical Interpretation	621
14.3.6	Convergence Aspects of the Fuzzy Clustering Algorithms	622
14.3.7	Alternating Cluster Estimation	622
14.4	Possibilistic Clustering	622
14.4.1	The Mode-Seeking Property	626
14.4.2	An Alternative Possibilistic Scheme	629
14.5	Hard Clustering Algorithms	629
14.5.1	The Isodata or k-Means or c-Means Algorithm	631
14.5.2	k-Medoids Algorithms	635
14.6	Vector Quantization	639
CHAPTER 15	CLUSTERING ALGORITHMS IV	653
15.1	Introduction	653
15.2	Clustering Algorithms Based on Graph Theory	653
15.2.1	Minimum Spanning Tree Algorithms	654
15.2.2	Algorithms Based on Regions of Influence	657
15.2.3	Algorithms Based on Directed Trees	658
15.3	Competitive Learning Algorithms	660
15.3.1	Basic Competitive Learning Algorithm	662
15.3.2	Leaky Learning Algorithm	664
15.3.3	Conscientious Competitive Learning Algorithms	665
15.3.4	Competitive Learning-Like Algorithms Associated with Cost Functions	666
15.3.5	Self-Organizing Maps	667
15.3.6	Supervised Learning Vector Quantization	668
15.4	Binary Morphology Clustering Algorithms (BMCAs)	669
15.4.1	Discretization	669
15.4.2	Morphological Operations	670
15.4.3	Determination of the Clusters in a Discrete Binary Set	673

15.4.4	Assignment of Feature Vectors to Clusters	675
15.4.5	The Algorithmic Scheme	675
15.5	Boundary Detection Algorithms	678
15.6	Valley-Seeking Clustering Algorithms	681
15.7	Clustering Via Cost Optimization (Revisited)	683
15.7.1	Branch and Bound Clustering Algorithms	683
15.7.2	Simulated Annealing	687
15.7.3	Deterministic Annealing	688
15.7.4	Clustering Using Genetic Algorithms	690
15.8	Kernel Clustering Methods	692
15.9	Density-Based Algorithms for Large Data Sets	695
15.9.1	The DBSCAN Algorithm	696
15.9.2	The DBCLASD Algorithm	699
15.9.3	The DENCLUE Algorithm	700
15.10	Clustering Algorithms for High-Dimensional Data Sets	702
15.10.1	Dimensionality Reduction Clustering Approach	703
15.10.2	Subspace Clustering Approach	705
15.11	Other Clustering Algorithms	718
CHAPTER 16	CLUSTER VALIDITY	733
16.1	Introduction	733
16.2	Hypothesis Testing Revisited	734
16.3	Hypothesis Testing in Cluster Validity	736
16.3.1	External Criteria	738
16.3.2	Internal Criteria	744
16.4	Relative Criteria	747
16.4.1	Hard Clustering	750
16.4.2	Fuzzy Clustering	756
16.5	Validity of Individual Clusters	763
16.5.1	External Criteria	763
16.5.2	Internal Criteria	764
16.6	Clustering Tendency	766
16.6.1	Tests for Spatial Randomness	770
Appendix A		
Hints from Probability and Statistics		785
Appendix B		
Linear Algebra Basics		797

Appendix C	
Cost Function Optimization	801
Appendix D	
Basic Definitions from Linear Systems Theory	819
Index	823