

Pattern Recognition

FOURTH EDITION

Sergios Theodoridis
Konstantinos Koutroumbas

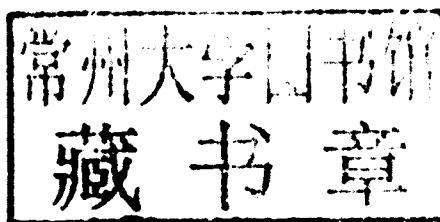


Pattern Recognition

Fourth Edition

Sergios Theodoridis

Konstantinos Koutroumbas



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. (∞)

Copyright © 2009, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, E-mail: permissions@elsevier.com. You may also complete your request on-line via the Elsevier homepage (<http://elsevier.com>), by selecting "Support & Contact" then "Copyright and Permission" and then "Obtaining Permissions."

Library of Congress Cataloging-in-Publication Data

Application submitted

British Library Cataloguing-in-Publication Data

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

ISBN: 978-1-59749-272-0

For information on all Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in Canada

09 10 11 12 13 14 15 16 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Pattern Recognition

Fourth Edition

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 Stamatoaponlos, 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	xv
CHAPTER 1	Introduction
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
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
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 l -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 <i>t</i> -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

Introduction

1.1 IS PATTERN RECOGNITION IMPORTANT?

Pattern recognition is the scientific discipline whose goal is the classification of *objects* into a number of categories or *classes*. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified. We will refer to these objects using the generic term *patterns*. Pattern recognition has a long history, but before the 1960s it was mostly the output of theoretical research in the area of statistics. As with everything else, the advent of computers increased the demand for practical applications of pattern recognition, which in turn set new demands for further theoretical developments. As our society evolves from the industrial to its postindustrial phase, automation in industrial production and the need for information handling and retrieval are becoming increasingly important. This trend has pushed pattern recognition to the high edge of today's engineering applications and research. Pattern recognition is an integral part of most *machine intelligence* systems built for decision making.

Machine vision is an area in which pattern recognition is of importance. A machine vision system captures images via a camera and analyzes them to produce descriptions of what is imaged. A typical application of a machine vision system is in the manufacturing industry, either for automated visual inspection or for automation in the assembly line. For example, in inspection, manufactured objects on a moving conveyor may pass the inspection station, where the camera stands, and it has to be ascertained whether there is a defect. Thus, images have to be analyzed online, and a pattern recognition system has to classify the objects into the "defect" or "nondefect" class. After that, an action has to be taken, such as to reject the offending parts. In an assembly line, different objects must be located and "recognized," that is, classified in one of a number of classes known *a priori*. Examples are the "screwdriver class," the "German key class," and so forth in a tools' manufacturing unit. Then a robot arm can move the objects in the right place.

Character (letter or number) recognition is another important area of pattern recognition, with major implications in automation and information handling. Optical character recognition (OCR) systems are already commercially available and more or less familiar to all of us. An OCR system has a "front-end" device consisting of a *light source*, a *scan lens*, a *document transport*, and a *detector*. At the output of