

# PATTERN RECOGNITION

SERGIOS THEODORIDIS

Departments of Informatics  
University of Athens  
Greece

*and*

KONSTANTINOS KOUTROUMBAS

The Greek Telecommunications Organization  
Athens, Greece



**ACADEMIC PRESS**

San Diego London Boston  
New York Sydney Tokyo Toronto

# CONTENTS

Preface	xiii
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 A Posteriori Probability Estimator	31
2.5.3 Bayesian Inference	33
2.5.4 Maximum Entropy Estimation	34
2.5.5 Mixture Models	35
2.5.6 Nonparametric Estimation	39
2.6 The Nearest Neighbor Rule	45
CHAPTER 3 LINEAR CLASSIFIERS	55
3.1 Introduction	55
3.2 Linear Discriminant Functions and Decision Hyperplanes	55
3.3 The Perceptron Algorithm	56
3.4 Least Squares Methods	65
3.4.1 Mean Square Error Estimation	65

3.4.2	Stochastic Approximation and the LMS Algorithm	68
3.4.3	Sum of Error Squares Estimation	70
3.5	Mean Square Estimation Revisited	72
3.5.1	Mean Square Error Regression	72
3.5.2	MSE Estimates Posterior Class Probabilities	73
3.5.3	The Bias–Variance Dilemma	76
CHAPTER 4 NONLINEAR CLASSIFIERS		81
4.1	Introduction	81
4.2	The XOR Problem	81
4.3	The Two-Layer Perceptron	83
4.3.1	Classification Capabilities of the Two-Layer Perceptron	86
4.4	Three-Layer Perceptrons	89
4.5	Algorithms Based on Correct Classification of the Training Set	90
4.6	The Backpropagation Algorithm	92
4.7	Variations on the Backpropagation Theme	100
4.8	The Cost Function Choice	103
4.9	Choice of the Network Size	106
4.10	A Simulation Example	112
4.11	Networks with Weight Sharing	114
4.12	Generalized Linear Classifiers	115
4.13	Capacity of the $l$ -Dimensional Space in Linear Dichotomies	117
4.14	Polynomial Classifiers	119
4.15	Radial Basis Function Networks	120
4.16	Universal Approximators	125
4.17	Decision Trees	126
CHAPTER 5 FEATURE SELECTION		139
5.1	Introduction	139
5.2	Preprocessing	140
5.2.1	Outlier Removal	140
5.2.2	Data Normalization	141
5.2.3	Missing Data	141
5.3	Feature Selection Based on Statistical Hypothesis Testing	142
5.3.1	Hypothesis Testing Basics	142
5.4	The Receiver Operating Characteristics Curve	149

5.5	Class Separability Measures	150
5.5.1	Divergence	150
5.5.2	Chernoff Bound and Brattacharyya Distance	153
5.5.3	Scatter Matrices	155
5.6	Feature Subset Selection	157
5.6.1	Scalar Feature Selection	157
5.6.2	Feature Vector Selection	159
5.7	Optimal Feature Generation	162
5.8	Neural Networks and Feature Generation/Selection	167
5.9	A Hint on the Vapnik–Chernovenkis Learning Theory	169
CHAPTER 6	FEATURE GENERATION I: LINEAR TRANSFORMS	181
6.1	Introduction	181
6.2	Basis Vectors and Images	182
6.3	The Karhunen–Loève Transform	184
6.4	The Singular Value Decomposition	189
6.5	The Discrete Fourier Transform (DFT)	192
6.5.1	One-Dimensional DFT	193
6.5.2	Two-Dimensional DFT	195
6.6	The Discrete Cosine and Sine Transforms	196
6.7	The Hadamard Transform	198
6.8	The Haar Transform	199
6.9	The Haar Expansion Revisited	200
6.10	Discrete Time Wavelet Transform (DTWT)	205
6.11	The Multiresolution Interpretation	215
6.12	Wavelet Packets	217
6.13	A Look at Two-Dimensional Generalizations	219
6.14	Applications	221
CHAPTER 7	FEATURE GENERATION II	233
7.1	Introduction	233
7.2	Regional Features	234
7.2.1	Features for Texture Characterization	234
7.2.2	Local Linear Transforms for Texture Feature Extraction	243
7.2.3	Moments	245
7.2.4	Parametric Models	249
7.3	Features for Shape and Size Characterization	258
7.3.1	Fourier Features	259
7.3.2	Chain Codes	262
7.3.3	Moment-Based Features	265

7.3.4	Geometric Features	266
7.4	A Glimpse at Fractals	267
7.4.1	Self-Similarity and Fractal Dimension	267
7.4.2	Fractional Brownian Motion	270
CHAPTER 8	TEMPLATE MATCHING	283
8.1	Introduction	283
8.2	Similarity Based on Optimal Path Searching Techniques	284
8.2.1	Bellman's Optimality Principle and Dynamic Programming	286
8.2.2	The Edit Distance	288
8.2.3	Dynamic Time Warping in Speech Recognition	291
8.3	Similarity Measures Based on Correlations	299
CHAPTER 9	CONTEXT-DEPENDENT CLASSIFICATION	307
9.1	Introduction	307
9.2	The Bayes Classifier	307
9.3	Markov Chain Models	308
9.4	The Viterbi Algorithm	309
9.5	Channel Equalization	312
9.6	Hidden Markov Models	316
9.7	A Discussion of Markov Random Fields	330
CHAPTER 10	SYSTEM EVALUATION	339
10.1	Introduction	339
10.2	Error Counting Approach	339
10.3	Exploiting the Finite Size of the Data Set	341
10.4	A Case Study from Medical Imaging	344
CHAPTER 11	CLUSTERING: BASIC CONCEPTS	351
11.1	Introduction	351
11.1.1	Applications of Cluster Analysis	354
11.1.2	Types of Features	355
11.1.3	Definitions of Clustering	356
11.2	Proximity Measures	358
11.2.1	Definitions	358
11.2.2	Proximity Measures between Two Points	361

	11.2.3	Proximity Functions between a Point and a Set	372
	11.2.4	Proximity Functions between Two Sets	377
CHAPTER 12	CLUSTERING ALGORITHMS I: SEQUENTIAL ALGORITHMS		383
	12.1	Introduction	383
	12.1.1	Number of Possible Clusterings	383
	12.2	Categories of Clustering Algorithms	385
	12.3	Sequential Clustering Algorithms	387
	12.3.1	Estimation of the Number of Clusters	389
	12.4	A Modification of BSAS	391
	12.5	A Two-Threshold Sequential Scheme	392
	12.6	Refinement Stages	395
	12.7	Neural Network Implementation	396
	12.7.1	Description of the Architecture	397
	12.7.2	Implementation of the BSAS Algorithm	398
CHAPTER 13	CLUSTERING ALGORITHMS II: HIERARCHICAL ALGORITHMS		403
	13.1	Introduction	403
	13.2	Agglomerative Algorithms	404
	13.2.1	Definition of Some Useful Quantities	405
	13.2.2	Agglomerative Algorithms Based on Matrix Theory	408
	13.2.3	Monotonicity and Crossover	415
	13.2.4	Implementational Issues	418
	13.2.5	Agglomerative Algorithms Based on Graph Theory	418
	13.2.6	Ties in the Proximity Matrix	427
	13.3	The Cophenetic Matrix	430
	13.4	Divisive Algorithms	431
	13.5	Choice of the Best Number of Clusters	433
CHAPTER 14	CLUSTERING ALGORITHMS III: SCHEMES BASED ON FUNCTION OPTIMIZATION		441
	14.1	Introduction	441
	14.2	Mixture Decomposition Schemes	443
	14.2.1	Compact and Hyperellipsoidal Clusters	445
	14.2.2	A Geometrical Interpretation	449

14.3	Fuzzy Clustering Algorithms	452
14.3.1	Point Representatives	456
14.3.2	Quadric Surfaces as Representatives	458
14.3.3	Hyperplane Representatives	468
14.3.4	Combining Quadric and Hyperplane Representatives	472
14.3.5	A Geometrical Interpretation	473
14.3.6	Convergence Aspects of the Fuzzy Clustering Algorithms	473
14.4	Possibilistic Clustering	473
14.4.1	The Mode-Seeking Property	478
14.4.2	An Alternative Possibilistic Scheme	479
14.5	Hard Clustering Algorithms	480
14.5.1	The Isodata or $k$ -Means or $c$ -Means Algorithm	482
14.6	Vector Quantization	484
CHAPTER 15	CLUSTERING ALGORITHMS IV	497
15.1	Introduction	497
15.2	Clustering Algorithms Based on Graph Theory	497
15.2.1	Minimum Spanning Tree Algorithms	498
15.2.2	Algorithms Based on Regions of Influence	501
15.2.3	Algorithms Based on Directed Trees	502
15.3	Competitive Learning Algorithms	504
15.3.1	Basic Competitive Learning Algorithms	506
15.3.2	Leaky Learning Algorithm	508
15.3.3	Conscientious Competitive Learning Algorithms	508
15.3.4	Competitive Learning-Like Algorithms Associated with Cost Functions	509
15.3.5	Self-Organizing Maps	510
15.3.6	Supervised Learning Vector Quantization	512
15.4	Branch and Bound Clustering Algorithms	512
15.5	Binary Morphology Clustering Algorithms (BMCAs)	516
15.5.1	Discretization	516
15.5.2	Morphological Operations	517
15.5.3	Determination of the Clusters in the Discrete Binary Set	520
15.5.4	Assignment of Feature Vectors to Clusters	522
15.5.5	The Algorithmic Scheme	523
15.6	Boundary Detection Algorithms	524

15.7	Valley-Seeking Clustering Algorithms	528
15.8	Clustering via Cost Optimization (Revisited)	530
	15.8.1 Simulated Annealing	530
	15.8.2 Deterministic Annealing	532
15.9	Clustering Using Genetic Algorithms	533
15.10	Other Clustering Algorithms	535
CHAPTER 16 CLUSTER VALIDITY		543
16.1	Introduction	543
16.2	Hypothesis Testing Revisited	544
16.3	Hypothesis Testing in Cluster Validity	546
	16.3.1 External Criteria	548
	16.3.2 Internal Criteria	553
16.4	Relative Criteria	557
	16.4.1 Hard Clustering	561
	16.4.2 Fuzzy Clustering	565
16.5	Validity of Individual Clusters	572
	16.5.1 External Criteria	572
	16.5.2 Internal Criteria	573
16.6	Clustering Tendency	574
	16.6.1 Tests for Spatial Randomness	578
Appendix A		
Hints from Probability and Statistics		593
Appendix B		
Linear Algebra Basics		601
Appendix C		
Cost Function Optimization		605
Appendix D		
Basic Definitions from Linear Systems Theory		613