

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction	1
1.2	Machine Recognition of Patterns: Preliminaries	3
1.2.1	Data Acquisition	4
1.2.2	Feature Selection	5
1.2.3	Classification	6
1.2.4	Clustering	8
1.3	Different Approaches	9
1.4	Connectionist Approach: Relevance and Features	11
1.5	Genetic Approach: Relevance and Features	13
1.6	Fuzzy Set-Theoretic Approach: Relevance and Features	14
1.7	Other Approaches	15
1.8	Applications of Pattern Recognition and Learning	16
1.9	Summary and Scope of the Book	17
<b>2</b>	<b>Genetic Algorithms</b>	<b>19</b>
2.1	Introduction	19
2.2	Traditional Versus Nontraditional Search	19
2.3	Overview of Genetic Algorithms	21
2.3.1	Basic Principles and Features	21
2.3.2	Encoding Strategy and Population	22
2.3.3	Evaluation	24
2.3.4	Genetic Operators	24
2.3.5	Parameters of Genetic Algorithms	27
2.3.6	Schema Theorem	27
2.4	Proof of Convergence of GAs	29
2.4.1	Markov Chain Modelling of GAs	29
2.4.2	Limiting Behavior of Elitist Model of GAs	31
2.5	Some Implementation Issues in GAs	35
2.6	Multiobjective Genetic Algorithms	40
2.7	Applications of Genetic Algorithms	46

2.8	Summary .....	51
<b>3</b>	<b>Supervised Classification Using Genetic Algorithms .....</b>	<b>53</b>
3.1	Introduction .....	53
3.2	Genetic Algorithms for Generating Fuzzy If-Then Rules .....	54
3.3	Genetic Algorithms and Decision Trees .....	57
3.4	<i>GA-classifier</i> : Genetic Algorithm for Generation of Class Boundaries .....	60
3.4.1	Principle of Hyperplane Fitting .....	61
3.4.2	Region Identification and Fitness Computation .....	62
3.4.3	Genetic Operations .....	65
3.5	Experimental Results .....	65
3.5.1	Results .....	69
3.5.2	Consideration of Higher-Order Surfaces .....	75
3.6	Summary .....	78
<b>4</b>	<b>Theoretical Analysis of the <i>GA-classifier</i> .....</b>	<b>81</b>
4.1	Introduction .....	81
4.2	Relationship with Bayes' Error Probability .....	82
4.3	Relationship Between $H_{opt}$ and $H_{GA}$ .....	88
4.3.1	Obtaining $H_{GA}$ from $H$ .....	88
4.3.2	How $H_{GA}$ Is Related to $H_{opt}$ .....	89
4.3.3	Some Points Related to $n$ and $H$ .....	89
4.4	Experimental Results .....	90
4.4.1	Data Sets .....	91
4.4.2	Learning the Class Boundaries and Performance on Test Data .....	93
4.4.3	Variation of Recognition Scores with $P_1$ .....	104
4.5	Summary .....	106
<b>5</b>	<b>Variable String Lengths in <i>GA-classifier</i> .....</b>	<b>109</b>
5.1	Introduction .....	109
5.2	Genetic Algorithm with Variable String Length and the Classification Criteria .....	110
5.3	Description of <i>VGA-Classifier</i> .....	111
5.3.1	Chromosome Representation and Population Initialization .....	111
5.3.2	Fitness Computation .....	113
5.3.3	Genetic Operators .....	114
5.4	Theoretical Study of <i>VGA-classifier</i> .....	117
5.4.1	Issues of Minimum <i>miss</i> and $H$ .....	117
5.4.2	Error Rate .....	118
5.5	Experimental Results .....	119
5.5.1	Data Sets .....	119
5.5.2	Results .....	120

5.6	<i>VGA-classifier</i> for the Design of a Multilayer Perceptron . . . . .	124
5.6.1	Analogy Between Multilayer Perceptron and <i>VGA-classifier</i> . . . . .	124
5.6.2	Deriving the MLP Architecture and the Connection Weights . . . . .	125
5.6.3	Postprocessing Step . . . . .	129
5.6.4	Experimental Results . . . . .	131
5.7	Summary . . . . .	132
<b>6</b>	<b>Chromosome Differentiation in <i>VGA-classifier</i></b> . . . . .	139
6.1	Introduction . . . . .	139
6.2	GACD: Incorporating Chromosome Differentiation in GA . . . . .	140
6.2.1	Motivation . . . . .	140
6.2.2	Description of GACD . . . . .	140
6.3	Schema Theorem for GACD . . . . .	143
6.3.1	Terminology . . . . .	143
6.3.2	Analysis of GACD . . . . .	143
6.4	<i>VGACD-classifier</i> : Incorporation of Chromosome Differentiation in <i>VGA-classifier</i> . . . . .	148
6.4.1	Population Initialization . . . . .	149
6.4.2	Fitness Computation and Genetic Operators . . . . .	150
6.5	Pixel Classification of Remotely Sensed Image . . . . .	150
6.5.1	Relevance of GA . . . . .	150
6.5.2	Experimental Results . . . . .	150
6.6	Summary . . . . .	154
<b>7</b>	<b>Multiobjective <i>VGA-classifier</i> and Quantitative Indices</b> . . . . .	159
7.1	Introduction . . . . .	159
7.2	Multiobjective Optimization . . . . .	160
7.3	Relevance of Multiobjective Optimization . . . . .	161
7.4	Multiobjective GA-Based Classifier . . . . .	162
7.4.1	Chromosome Representation . . . . .	162
7.4.2	Fitness Computation . . . . .	162
7.4.3	Selection . . . . .	163
7.4.4	Crossover . . . . .	164
7.4.5	Mutation . . . . .	164
7.4.6	Incorporating Elitism . . . . .	165
7.4.7	<i>PAES-classifier</i> : The Classifier Based on Pareto Archived Evolution Strategy . . . . .	167
7.5	Validation and Testing . . . . .	169
7.6	Indices for Comparing MO Solutions . . . . .	170
7.6.1	Measures Based on Position of Nondominated Front . . . . .	170
7.6.2	Measures Based on Diversity of the Solutions . . . . .	171
7.7	Experimental Results . . . . .	172
7.7.1	Parameter Values . . . . .	173

7.7.2	Comparison of Classification Performance	173
7.8	Summary	179
<b>8</b>	<b>Genetic Algorithms in Clustering</b>	<b>181</b>
8.1	Introduction	181
8.2	Basic Concepts and Preliminary Definitions	182
8.3	Clustering Algorithms	184
8.3.1	K-Means Clustering Algorithm	184
8.3.2	Single-Linkage Clustering Algorithm	185
8.3.3	Fuzzy c-Means Clustering Algorithm	186
8.4	Clustering Using GAs: Fixed Number of Crisp Clusters	187
8.4.1	Encoding Strategy	188
8.4.2	Population Initialization	188
8.4.3	Fitness Computation	188
8.4.4	Genetic Operators	189
8.4.5	Experimental Results	189
8.5	Clustering Using GAs: Variable Number of Crisp Clusters	192
8.5.1	Encoding Strategy and Population Initialization	192
8.5.2	Fitness Computation	193
8.5.3	Genetic Operators	193
8.5.4	Some Cluster Validity Indices	194
8.5.5	Experimental Results	196
8.6	Clustering Using GAs: Variable Number of Fuzzy Clusters	205
8.6.1	Fitness Computation	205
8.6.2	Experimental Results	206
8.7	Summary	212
<b>9</b>	<b>Genetic Learning in Bioinformatics</b>	<b>213</b>
9.1	Introduction	213
9.2	Bioinformatics: Concepts and Features	214
9.2.1	Basic Concepts of Cell Biology	214
9.2.2	Different Bioinformatics Tasks	216
9.3	Relevance of Genetic Algorithms in Bioinformatics	216
9.4	Bioinformatics Tasks and Application of GAs	220
9.4.1	Alignment and Comparison of DNA, RNA and Protein Sequences	220
9.4.2	Gene Mapping on Chromosomes	223
9.4.3	Gene Finding and Promoter Identification from DNA Sequences	224
9.4.4	Interpretation of Gene Expression and Microarray Data	226
9.4.5	Gene Regulatory Network Identification	227
9.4.6	Construction of Phylogenetic Trees for Studying Evolutionary Relationship	228
9.4.7	DNA Structure Prediction	229
9.4.8	RNA Structure Prediction	231

9.4.9	Protein Structure Prediction and Classification . . . . .	233
9.4.10	Molecular Design and Docking . . . . .	236
9.5	Experimental Results . . . . .	238
9.6	Summary . . . . .	239
<b>10</b>	<b>Genetic Algorithms and Web Intelligence . . . . .</b>	<b>243</b>
10.1	Introduction . . . . .	243
10.2	Web Mining . . . . .	244
10.2.1	Web Mining Components and Methodologies . . . . .	246
10.2.2	Web Mining Categories . . . . .	246
10.2.3	Challenges and Limitations in Web Mining . . . . .	248
10.3	Genetic Algorithms in Web Mining . . . . .	250
10.3.1	Search and Retrieval . . . . .	250
10.3.2	Query Optimization and Reformulation . . . . .	252
10.3.3	Document Representation and Personalization . . . . .	254
10.3.4	Distributed Mining . . . . .	254
10.4	Summary . . . . .	255
<b>A</b>	<b><math>\epsilon</math>-Optimal Stopping Time for GAs . . . . .</b>	<b>257</b>
A.1	Introduction . . . . .	257
A.2	Foundation . . . . .	257
A.3	Fitness Function . . . . .	259
A.4	Upper Bound for Optimal Stopping Time . . . . .	261
A.5	Mutation Probability and $\epsilon$ -Optimal Stopping Time . . . . .	264
<b>B</b>	<b>Data Sets Used for the Experiments . . . . .</b>	<b>269</b>
<b>C</b>	<b>Variation of Error Probability with <math>P_1</math> . . . . .</b>	<b>275</b>
	<b>References . . . . .</b>	<b>277</b>
	<b>Index . . . . .</b>	<b>309</b>



Classification and Learning Using Genetic Algorithms

Applications in Bioinformatics and Web Intelligence

Bandyopadhyay, S.; Pal, S.K.

2007, XVI, 311 p. 87 illus., Hardcover

ISBN: 978-3-540-49606-9