CONTENTS

Series Editor's Foreword xiii

Preface xv

Notation xix

Chapter 1 Introduction

1.1 The Information-Handling Problem 1
1.2 Basic Concepts of Pattern Recognition 5
1.3 Fundamental Problems in Pattern Recognition System Design 9
1.4 Design Concepts and Methodologies 17
1.5 Examples of Automatic Pattern Recognition Systems . . 21
1.6 A Simple Automatic Pattern Recognition Model 36

Chapter 2 Decision Functions

2.1 Introduction 39
2.2 Linear Decision Functions 40
2.3 Generalized Decision Functions 48
2.4 Pattern Space and Weight Space 53
2.5 Geometrical Properties 55
2.5.1 Hyperplane Properties 55
2.5.2 Dichotomies 58
2.5.3 Dichotomization Capacity of Generalized Decision Functions 60
2.6 Implementation of Decision Functions 62
2.7 Functions of Several Variables 64
2.7.1 Definitions 65

2.7.2 Construction of Multivariate Functions 67
2.7.3 Orthogonal and Orthonormal Systems of Functions 68
2.8 Concluding Remarks 72
References 73
Problems 73

Chapter 3 Pattern Classification by Distance Functions

3.1 Introduction 75
3.2 Minimum-Distance Pattern Classification 76
3.2.1 Single Prototypes 77
3.2.2 Multiprototypes 78
3.2.3 Extension of Minimum-Distance Classification Concepts 81
3.2.4 A Design Example 83
3.3 Cluster Seeking 86
3.3.1 Measures of Similarity 87
3.3.2 Clustering Criteria 89
3.3.3 A Simple Cluster-Seeking Algorithm 90
3.3.4 Maximin-Distance Algorithm 92
3.3.5 *K*-Means Algorithm 94
3.3.6 Isodata Algorithm 97
3.3.7 Evaluation of Clustering Results 104
3.3.8 Graph-Theoretic Approach 106
3.4 Unsupervised Pattern Recognition 107
3.5 Concluding Remarks 108
References 108
Problems 109

Chapter 4 Pattern Classification by Likelihood Functions

4.1 Introduction 110
4.2 Pattern Classification as a Statistical Decision Problem . . 111
4.3 Bayes Classifier for Normal Patterns 119
4.4 Error Probabilities 124
4.5 A Family of Important Probability Density Functions . . 130
4.6 Estimation of Probability Density Functions 134
4.6.1 Form of the Probability Density Function 134
4.6.2 Estimation of the Mean Vector and Covariance Matrix 137
4.6.3 Bayesian Learning of the Mean Vector and Covariance Matrix 139

4.6.4 Functional Approximation of Probability Density Functions . . . 145
4.7 Concluding Remarks . . . 154
References . . . 155
Problems . . . 156

Chapter 5 Trainable Pattern Classifiers—The Deterministic Approach

5.1 Introduction . . . 158
5.2 The Perceptron Approach . . . 159
5.2.1 The Reward-Punishment Concept . . . 161
5.2.2 Proof of Convergence . . . 165
5.2.3 Variations of the Perceptron Approach . . . 168
5.3 Derivation of Pattern Classification Algorithms . . . 169
5.3.1 The Gradient Technique . . . 169
5.3.2 Perceptron Algorithm . . . 171
5.3.3 A Least-Mean-Square-Error Algorithm . . . 173
5.3.4 Convergence Proof of the LMSE Algorithm . . . 178
5.4 Multicategory Classification . . . 181
5.5 Learning and Generalization . . . 186
5.6 The Potential Function Approach . . . 187
5.6.1 Generation of Decision Functions . . . 188
5.6.2 Selection of Potential Functions . . . 193
5.6.3 Geometrical Interpretation and Weight Adjustment 202
5.6.4 Convergence of Training Algorithms . . . 208
5.6.5 Multiclass Generalization . . . 212
5.7 Concluding Remarks . . . 213
References . . . 214
Problems . . . 215

Chapter 6 Trainable Pattern Classifiers—The Statistical Approach

6.1 Introduction . . . 217
6.2 Stochastic Approximation Methods . . . 218
6.2.1 The Robbins-Monro Algorithm . . . 219
6.2.2 Speed of Convergence . . . 224
6.2.3 Multidimensional Extension . . . 225
6.3 Derivation of Pattern Classification Algorithms . . . 226
6.3.1 Estimation of Optimum Decision Functions by Stochastic Approximation Methods . . . 226
6.3.2 Increment-Correction Algorithm . . . 229

6.3.3 Least-Mean-Square-Error Algorithm 233
6.4 The Method of Potential Functions 235
6.5 Concluding Remarks 239
References 240
Problems 241

Chapter 7 Pattern Preprocessing and Feature Selection

7.1 Introduction 243
7.2 Distance Measures 247
7.3 Clustering Transformations and Feature Ordering 250
7.4 Clustering in Feature Selection 259
7.5 Feature Selection Through Entropy Minimization 263
7.6 Feature Selection Through Orthogonal Expansions. . . . 269
7.6.1 Review of the Fourier Series Expansion 269
7.6.2 Karhunen-Loève Expansion 271
7.7 Feature Selection Through Functional Approximation . . 283
7.7.1 Functional Expansion 283
7.7.2 Stochastic Approximation Formulation. 288
7.7.3 Kernel Approximation 290
7.7.4 Use of Feature Functions in Classification 291
7.8 Divergence Concept 291
7.9 Feature Selection Through Divergence Maximization . . . 298
7.10 Binary Feature Selection 307
7.10.1 A Sequential Algorithm 308
7.10.2 A Parallel Algorithm. 310
7.11 Concluding Remarks 313
References 314
Problems 314

Chapter 8 Syntactic Pattern Recognition

8.1 Introduction 316
8.2 Concepts From Formal Language Theory 317
8.2.1 Definitions 317
8.2.2 Types of Grammars 320
8.3 Formulation of the Syntactic Pattern Recognition Problem 323
8.4 Syntactic Pattern Description 324
8.5 Recognition Grammars 328
8.5.1 Syntax-Directed Recognition 328
8.5.2 Recognition of Graph-Like Patterns 331
8.5.3 Recognition of Tree Structures 339

8.6 Statistical Considerations 340
8.6.1 Stochastic Grammars and Languages 341
8.6.2 Learning the Production Probabilities 343
8.7 Learning and Grammatical Inference 350
8.7.1 Inference of String Grammars 350
8.7.2 Inference of Two-Dimensional Grammars 355
8.8 Automata as Pattern Recognizers 357
8.9 Concluding Remarks 360
References 361
Problems 361

Bibliography 363

Index 372