Introduction to Pattern Recognition
Introduction to Pattern Recognition

A MATLAB® Approach

Sergios Theodoridis
Aggelos Pikrakis
Konstantinos Koutroumbas
Dionisis Cavouras
# Contents

**Preface** ........................................................................................................................................... ix

**CHAPTER 1**  **Classifiers Based on Bayes Decision Theory**  .................................................................................. 1

1.1 Introduction .................................................................................................................................................. 1

1.2 Bayes Decision Theory .................................................................................................................................. 1

1.3 The Gaussian Probability Density Function ............................................................................................ 2

1.4 Minimum Distance Classifiers .................................................................................................................. 6

1.4.1 The Euclidean Distance Classifier ....................................................................................................... 6

1.4.2 The Mahalanobis Distance Classifier .................................................................................................... 6

1.4.3 Maximum Likelihood Parameter Estimation of Gaussian pdfs ............................................................ 7

1.5 Mixture Models ........................................................................................................................................... 11

1.6 The Expectation-Maximization Algorithm ............................................................................................... 13

1.7 Parzen Windows .......................................................................................................................................... 19

1.8 k-Nearest Neighbor Density Estimation ..................................................................................................... 21

1.9 The Naive Bayes Classifier ........................................................................................................................ 22

1.10 The Nearest Neighbor Rule ..................................................................................................................... 25

**CHAPTER 2**  **Classifiers Based on Cost Function Optimization** .............................................................. 29

2.1 Introduction ................................................................................................................................................. 29

2.2 The Perceptron Algorithm ......................................................................................................................... 30

2.2.1 The Online Form of the Perceptron Algorithm ................................................................................... 33

2.3 The Sum of Error Squares Classifier ....................................................................................................... 35

2.3.1 The Multiclass LS Classifier ............................................................................................................... 39

2.4 Support Vector Machines: The Linear Case ............................................................................................... 43

2.4.1 Multiclass Generalizations .................................................................................................................. 48

2.5 SVM: The Nonlinear Case .......................................................................................................................... 50

2.6 The Kernel Perceptron Algorithm ............................................................................................................. 58

2.7 The AdaBoost Algorithm .......................................................................................................................... 63

2.8 Multilayer Perceptrons ............................................................................................................................... 66

**CHAPTER 3**  **Data Transformation: Feature Generation and Dimensionality Reduction** .............................. 79

3.1 Introduction ................................................................................................................................................. 79

3.2 Principal Component Analysis ................................................................................................................... 79

3.3 The Singular Value Decomposition Method ............................................................................................... 84

3.4 Fisher’s Linear Discriminant Analysis ....................................................................................................... 87
### Contents

**3.5** The Kernel PCA ............................................................. 92
**3.6** Laplacian Eigenmap ....................................................... 101

**CHAPTER 4** Feature Selection ............................................... 107
**4.1** Introduction .................................................................. 107
**4.2** Outlier Removal ........................................................... 107
**4.3** Data Normalization ....................................................... 108
**4.4** Hypothesis Testing: The \( t \)-Test ................................. 111
**4.5** The Receiver Operating Characteristic Curve ............. 113
**4.6** Fisher’s Discriminant Ratio .......................................... 114
**4.7** Class Separability Measures ......................................... 117
   - 4.7.1 Divergence ......................................................... 118
   - 4.7.2 Bhattacharyya Distance and Chernoff Bound .......... 119
   - 4.7.3 Measures Based on Scatter Matrices ................. 120
**4.8** Feature Subset Selection ............................................... 122
   - 4.8.1 Scalar Feature Selection .................................... 123
   - 4.8.2 Feature Vector Selection .................................... 124

**CHAPTER 5** Template Matching ............................................. 137
**5.1** Introduction .................................................................. 137
**5.2** The Edit Distance ........................................................ 137
**5.3** Matching Sequences of Real Numbers ....................... 139
**5.4** Dynamic Time Warping in Speech Recognition .......... 143

**CHAPTER 6** Hidden Markov Models ..................................... 147
**6.1** Introduction .................................................................. 147
**6.2** Modeling ..................................................................... 147
**6.3** Recognition and Training ............................................. 148

**CHAPTER 7** Clustering ......................................................... 159
**7.1** Introduction .................................................................. 159
**7.2** Basic Concepts and Definitions ................................... 159
**7.3** Clustering Algorithms .................................................. 160
**7.4** Sequential Algorithms .................................................. 161
   - 7.4.1 BSAS Algorithm .............................................. 161
   - 7.4.2 Clustering Refinement ....................................... 162
**7.5** Cost Function Optimization Clustering Algorithms .... 168
   - 7.5.1 Hard Clustering Algorithms ............................. 168
   - 7.5.2 Nonhard Clustering Algorithms ....................... 184
**7.6** Miscellaneous Clustering Algorithms .......................... 189
7.7 Hierarchical Clustering Algorithms .......................................................... 198
    7.7.1 Generalized Agglomerative Scheme ............................................... 199
    7.7.2 Specific Agglomerative Clustering Algorithms ............................... 200
    7.7.3 Choosing the Best Clustering .......................................................... 203

Appendix ........................................................................................................... 209

References ......................................................................................................... 215

Index ................................................................................................................. 217
Preface

The aim of this book is to serve pedagogic goals as a complement of the book *Pattern Recognition*, 4th Edition, by S. Theodoridis and K. Koutroumbas (Academic Press, 2009). It is the offspring of our experience in teaching pattern recognition for a number of years to different audiences such as students with good enough mathematical background, students who are more practice-oriented, professional engineers, and computer scientists attending short training courses. The book unravels along two directions.

The first is to develop a set of MATLAB-based examples so that students will be able to experiment with methods and algorithms met in the various stages of designing a pattern recognition system—that is, classifier design, feature selection and generation, and system evaluation. To this end, we have made an effort to “design” examples that will help the reader grasp the basics behind each method as well as the respective cons and pros. In pattern recognition, there are no magic recipes that dictate which method or technique to use for a specific problem. Very often, old good and simple (in concept) techniques can compete, from an efficiency point of view, with more modern and elaborate techniques. To this end, that is, selecting the most appropriate technique, it is unfortunate that these days more and more people follow the so-called black-box approach: try different techniques, using a related S/W package to play with a set of parameters, even if the real meaning of these parameters is not really understood.

Such an “unscientific” approach, which really prevents thought and creation, also deprives the “user” of the ability to understand, explain, and interpret the obtained results. For this reason, most of the examples in this book use simulated data. Hence, one can experiment with different parameters and study the behavior of the respective method/algorith. Having control of the data, readers will be able to “study,” “investigate,” and get familiar with the pros and cons of a technique. One can create data that can push a technique to its “limits”—that is, where it fails. In addition, most of the real-life problems are solved in high-dimensional spaces, where visualization is impossible; yet, visualizing geometry is one of the most powerful and effective pedagogic tools so that a newcomer to the field will be able to “see” how various methods work. The 3-dimensional space is one of the most primitive and deep-rooted experiences in the human brain because everyone is acquainted with and has a good feeling about and understanding of it.

The second direction is to provide a summary of the related theory, without mathematics. We have made an effort, as much as possible, to present the theory using arguments based on physical reasoning, as well as point out the role of the various parameters and how they can influence the performance of a method/algorith. Nevertheless, for a more thorough understanding, the mathematical formulation cannot be bypassed. It is “there” where the real worthy secrets of a method are, where the deep understanding has its undisputable roots and grounds, where science lies. Theory and practice are interrelated—one cannot be developed without the other. This is the reason that we consider this book a complement of the previously published one. We consider it another branch leaning toward the practical side, the other branch being the more theoretical one. Both branches are necessary to form the pattern-recognition tree, which has its roots in the work of hundreds of researchers who have effortlessly contributed, over a number of decades, both in theory and practice.
All the MATLAB functions used throughout this book can be downloaded from the companion website for this book at www.elsevierdirect.com/9780123744869. Note that, when running the MATLAB code in the book, the results may slightly vary among different versions of MATLAB. Moreover, we have made an effort to minimize dependencies on MATLAB toolboxes, as much as possible, and have developed our own code.

Also, in spite of the careful proofreading of the book, it is still possible that some typos may have escaped. The authors would appreciate readers notifying them of any that are found, as well as suggestions related to the MATLAB code.
Introduction to Pattern Recognition