# Contents

**Acknowledgements**

**Preface**

## 1 Introduction
1.1 Past, Present and Future 1
1.2 About this Book 9
Bibliography 12

## 2 Case Studies
2.1 Introduction 15
2.2 Datasets, Matrices and Vectors 17
2.3 Case Study 1: Forensic Analysis of Banknotes 20
2.4 Case Study 2: Near Infrared Spectroscopic Analysis of Food 23
2.5 Case Study 3: Thermal Analysis of Polymers 25
2.6 Case Study 4: Environmental Pollution using Headspace Mass Spectrometry 27
2.7 Case Study 5: Human Sweat Analysed by Gas Chromatography Mass Spectrometry 30
2.8 Case Study 6: Liquid Chromatography Mass Spectrometry of Pharmaceutical Tablets 32
2.9 Case Study 7: Atomic Spectroscopy for the Study of Hypertension 34
2.10 Case Study 8: Metabolic Profiling of Mouse Urine by Gas Chromatography of Urine Extracts 36
2.11 Case Study 9: Nuclear Magnetic Resonance Spectroscopy for Salival Analysis of the Effect of Mouthwash 37
2.12 Case Study 10: Simulations 38
2.13 Case Study 11: Null Dataset 40
2.14 Case Study 12: GCMS and Microbiology of Mouse Scent Marks 42
Bibliography 45

## 3 Exploratory Data Analysis
3.1 Introduction 47
3.2 Principal Components Analysis 49
3.2.1 Background 49
3.2.2 Scores and Loadings 50
3.2.3 Eigenvalues 53
3.2.4 PCA Algorithm 57
3.2.5 Graphical Representation 57

3.3 Dissimilarity Indices, Principal Co-ordinates Analysis and Ranking 75
  3.3.1 Dissimilarity 75
  3.3.2 Principal Co-ordinates Analysis 80
  3.3.3 Ranking 84

3.4 Self Organizing Maps 87
  3.4.1 Background 87
  3.4.2 SOM Algorithm 88
  3.4.3 Initialization 89
  3.4.4 Training 90
  3.4.5 Map Quality 93
  3.4.6 Visualization 95

Bibliography 105

4 Preprocessing 107
  4.1 Introduction 107
  4.2 Data Scaling 108
    4.2.1 Transforming Individual Elements 108
    4.2.2 Row Scaling 117
    4.2.3 Column Scaling 124
  4.3 Multivariate Methods of Data Reduction 129
    4.3.1 Largest Principal Components 129
    4.3.2 Discriminatory Principal Components 137
    4.3.3 Partial Least Squares Discriminatory Analysis Scores 145
  4.4 Strategies for Data Preprocessing 150
    4.4.1 Flow Charts 150
    4.4.2 Level 1 153
    4.4.3 Level 2 161
    4.4.4 Level 3 162
    4.4.5 Level 4 175

Bibliography 176

5 Two Class Classifiers 177
  5.1 Introduction 177
    5.1.1 Two Class Classifiers 178
    5.1.2 Preprocessing 180
    5.1.3 Notation 180
    5.1.4 Autoprediction and Class Boundaries 181
  5.2 Euclidean Distance to Centroids 184
  5.3 Linear Discriminant Analysis 185
  5.4 Quadratic Discriminant Analysis 192
  5.5 Partial Least Squares Discriminant Analysis 196
11 Class Separation Indices 469
11.1 Introduction 469
11.2 Davies Bouldin Index 470
11.3 Silhouette Width and Modified Silhouette Width 475
  11.3.1 Silhouette Width 475
  11.3.2 Modified Silhouette Width 475
11.4 Overlap Coefficient 477
Bibliography 478

12 Comparing Different Patterns 479
12.1 Introduction 479
12.2 Correlation Based Methods 481
  12.2.1 Mantel Test 481
  12.2.2 $R_V$ Coefficient 483
12.3 Consensus PCA 484
12.4 Procrustes Analysis 487
Bibliography 492

Index 493