Contents

List of Algorithms *page* xiii
Notation xv
Preface xix

I CLASSICAL METHODS

1 Multidimensional Data 3
1.1 Multivariate and High-Dimensional Problems 3
1.2 Visualisation 4
1.2.1 Three-Dimensional Visualisation 4
1.2.2 Parallel Coordinate Plots 6
1.3 Multivariate Random Vectors and Data 8
1.3.1 The Population Case 8
1.3.2 The Random Sample Case 9
1.4 Gaussian Random Vectors 11
1.4.1 The Multivariate Normal Distribution and the Maximum Likelihood Estimator 11
1.4.2 Marginal and Conditional Normal Distributions 13
1.5 Similarity, Spectral and Singular Value Decomposition 14
1.5.1 Similar Matrices 14
1.5.2 Spectral Decomposition for the Population Case 14
1.5.3 Decompositions for the Sample Case 16

2 Principal Component Analysis 18
2.1 Introduction 18
2.2 Population Principal Components 19
2.3 Sample Principal Components 22
2.4 Visualising Principal Components 27
2.4.1 Scree, Eigenvalue and Variance Plots 27
2.4.2 Two- and Three-Dimensional PC Score Plots 30
2.4.3 Projection Plots and Estimates of the Density of the Scores 31
2.5 Properties of Principal Components 34
2.5.1 Correlation Structure of $\mathbf{X}$ and Its PCs 34
2.5.2 Optimality Properties of PCs 37

2.6 Standardised Data and High-Dimensional Data 42
2.6.1 Scaled and Sphered Data 42
2.6.2 High-Dimensional Data 47
2.7 Asymptotic Results 55
2.7.1 Classical Theory: Fixed Dimension d 55
2.7.2 Asymptotic Results when d Grows 57
2.8 Principal Component Analysis, the Number of Components and Regression 62
2.8.1 Number of Principal Components Based on the Likelihood 62
2.8.2 Principal Component Regression 65

3 Canonical Correlation Analysis 70
3.1 Introduction 70
3.2 Population Canonical Correlations 71
3.3 Sample Canonical Correlations 76
3.4 Properties of Canonical Correlations 82
3.5 Canonical Correlations and Transformed Data 88
3.5.1 Linear Transformations and Canonical Correlations 88
3.5.2 Transforms with Non-Singular Matrices 90
3.5.3 Canonical Correlations for Scaled Data 98
3.5.4 Maximum Covariance Analysis 100
3.6 Asymptotic Considerations and Tests for Correlation 100
3.7 Canonical Correlations and Regression 104
3.7.1 The Canonical Correlation Matrix in Regression 105
3.7.2 Canonical Correlation Regression 108
3.7.3 Partial Least Squares 109
3.7.4 The Generalised Eigenvalue Problem 113

4 Discriminant Analysis 116
4.1 Introduction 116
4.2 Classes, Labels, Rules and Decision Functions 118
4.3 Linear Discriminant Rules 120
4.3.1 Fisher's Discriminant Rule for the Population 120
4.3.2 Fisher's Discriminant Rule for the Sample 123
4.3.3 Linear Discrimination for Two Normal Populations or Classes 127
4.4 Evaluation of Rules and Probability of Misclassification 129
4.4.1 Boundaries and Discriminant Regions 129
4.4.2 Evaluation of Discriminant Rules 131
4.5 Discrimination under Gaussian Assumptions 136
4.5.1 Two and More Normal Classes 136
4.5.2 Gaussian Quadratic Discriminant Analysis 140
4.6 Bayesian Discrimination 143
4.6.1 Bayes Discriminant Rule 143
4.6.2 Loss and Bayes Risk 146
4.7 Non-Linear, Non-Parametric and Regularised Rules 148
4.7.1 Nearest-Neighbour Discrimination 149

4.7.2 Logistic Regression and Discrimination 153
4.7.3 Regularised Discriminant Rules 154
4.7.4 Support Vector Machines 155
4.8 Principal Component Analysis, Discrimination and Regression 157
4.8.1 Discriminant Analysis and Linear Regression 157
4.8.2 Principal Component Discriminant Analysis 158
4.8.3 Variable Ranking for Discriminant Analysis 159

Problems for Part I 165

II FACTORS AND GROUPINGS

5 Norms, Proximities, Features and Dualities 175
5.1 Introduction 175
5.2 Vector and Matrix Norms 176
5.3 Measures of Proximity 176
5.3.1 Distances 176
5.3.2 Dissimilarities 178
5.3.3 Similarities 179
5.4 Features and Feature Maps 180
5.5 Dualities for $\mathbb{X}$ and $\mathbb{X}^T$ 181

6 Cluster Analysis 183
6.1 Introduction 183
6.2 Hierarchical Agglomerative Clustering 185
6.3 k-Means Clustering 191
6.4 Second-Order Polynomial Histogram Estimators 199
6.5 Principal Components and Cluster Analysis 207
6.5.1 k-Means Clustering for Principal Component Data 207
6.5.2 Binary Clustering of Principal Component Scores and Variables 210
6.5.3 Clustering High-Dimensional Binary Data 212
6.6 Number of Clusters 216
6.6.1 Quotients of Variability Measures 216
6.6.2 The Gap Statistic 217
6.6.3 The Prediction Strength Approach 219
6.6.4 Comparison of $\hat{k}$-Statistics 220

7 Factor Analysis 223
7.1 Introduction 223
7.2 Population k-Factor Model 224
7.3 Sample k-Factor Model 227
7.4 Factor Loadings 228
7.4.1 Principal Components and Factor Analysis 228
7.4.2 Maximum Likelihood and Gaussian Factors 233
7.5 Asymptotic Results and the Number of Factors 236
7.6 Factor Scores and Regression 239

7.6.1 Principal Component Factor Scores 239
7.6.2 Bartlett and Thompson Factor Scores 240
7.6.3 Canonical Correlations and Factor Scores 241
7.6.4 Regression-Based Factor Scores 242
7.6.5 Factor Scores in Practice 244
7.7 Principal Components, Factor Analysis and Beyond 245

8 Multidimensional Scaling 248
8.1 Introduction 248
8.2 Classical Scaling 249
8.2.1 Classical Scaling and Principal Coordinates 251
8.2.2 Classical Scaling with $\mathcal{S}$train 254
8.3 Metric Scaling 257
8.3.1 Metric Dissimilarities and Metric Stresses 258
8.3.2 Metric $\mathcal{S}$train 261
8.4 Non-Metric Scaling 263
8.4.1 Non-Metric Stress and the Shepard Diagram 263
8.4.2 Non-Metric $\mathcal{S}$train 268
8.5 Data and Their Configurations 268
8.5.1 HDLSS Data and the $\mathbb{X}$ and $\mathbb{X}^T$ Duality 269
8.5.2 Procrustes Rotations 271
8.5.3 Individual Differences Scaling 273
8.6 Scaling for Grouped and Count Data 274
8.6.1 Correspondence Analysis 274
8.6.2 Analysis of Distance 279
8.6.3 Low-Dimensional Embeddings 282

Problems for Part II 286

III NON-GAUSSIAN ANALYSIS

9 Towards Non-Gaussianity 295
9.1 Introduction 295
9.2 Gaussianity and Independence 296
9.3 Skewness, Kurtosis and Cumulants 297
9.4 Entropy and Mutual Information 299
9.5 Training, Testing and Cross-Validation 301
9.5.1 Rules and Prediction 302
9.5.2 Evaluating Rules with the Cross-Validation Error 302

10 Independent Component Analysis 305
10.1 Introduction 305
10.2 Sources and Signals 307
10.2.1 Population Independent Components 307
10.2.2 Sample Independent Components 308
10.3 Identification of the Sources 310
10.4 Mutual Information and Gaussianity 314

10.4.1 Independence, Uncorrelatedness and Non-Gaussianity 314
10.4.2 Approximations to the Mutual Information 317
10.5 Estimation of the Mixing Matrix 320
10.5.1 An Estimating Function Approach 321
10.5.2 Properties of Estimating Functions 322
10.6 Non-Gaussianity and Independence in Practice 324
10.6.1 Independent Component Scores and Solutions 324
10.6.2 Independent Component Solutions for Real Data 326
10.6.3 Performance of $\widehat{\mathfrak{I}}$ for Simulated Data 331
10.7 Low-Dimensional Projections of High-Dimensional Data 335
10.7.1 Dimension Reduction and Independent Component Scores 335
10.7.2 Properties of Low-Dimensional Projections 339
10.8 Dimension Selection with Independent Components 343

11 Projection Pursuit 349
11.1 Introduction 349
11.2 One-Dimensional Projections and Their Indices 350
11.2.1 Population Projection Pursuit 350
11.2.2 Sample Projection Pursuit 356
11.3 Projection Pursuit with Two- and Three-Dimensional Projections 359
11.3.1 Two-Dimensional Indices: $\mathcal{Q}_{\mathrm{E}}$, $\mathcal{Q}_{\mathrm{C}}$ and $\mathcal{Q}_{\mathrm{U}}$ 359
11.3.2 Bivariate Extension by Removal of Structure 361
11.3.3 A Three-Dimensional Cumulant Index 363
11.4 Projection Pursuit in Practice 363
11.4.1 Comparison of Projection Pursuit and Independent Component Analysis 364
11.4.2 From a Cumulant-Based Index to FastICA Scores 365
11.4.3 The Removal of Structure and FastICA 366
11.4.4 Projection Pursuit: A Continuing Pursuit 371
11.5 Theoretical Developments 373
11.5.1 Theory Relating to $\mathcal{Q}_{\mathrm{R}}$ 373
11.5.2 Theory Relating to $\boldsymbol{\mathcal{Q}}_{\boldsymbol{U}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{D}}$ 374
11.6 Projection Pursuit Density Estimation and Regression 376
11.6.1 Projection Pursuit Density Estimation 376
11.6.2 Projection Pursuit Regression 378

12 Kernel and More Independent Component Methods 381
12.1 Introduction 381
12.2 Kernel Component Analysis 382
12.2.1 Feature Spaces and Kernels 383
12.2.2 Kernel Principal Component Analysis 385
12.2.3 Kernel Canonical Correlation Analysis 389
12.3 Kernel Independent Component Analysis 392
12.3.1 The $\mathcal{F}$-Correlation and Independence 392
12.3.2 Estimating the $\mathcal{F}$-Correlation 394

12.3.3 Comparison of Non-Gaussian and Kernel Independent Components Approaches 396
12.4 Independent Components from Scatter Matrices (aka Invariant Coordinate Selection) 402
12.4.1 Scatter Matrices 403
12.4.2 Population Independent Components from Scatter Matrices 404
12.4.3 Sample Independent Components from Scatter Matrices 407
12.5 Non-Parametric Estimation of Independence Criteria 413
12.5.1 A Characteristic Function View of Independence 413
12.5.2 An Entropy Estimator Based on Order Statistics 416
12.5.3 Kernel Density Estimation of the Unmixing Matrix 417

13 Feature Selection and Principal Component Analysis Revisited 421
13.1 Introduction 421
13.2 Independent Components and Feature Selection 423
13.2.1 Feature Selection in Supervised Learning 423
13.2.2 Best Features and Unsupervised Decisions 426
13.2.3 Test of Gaussianity 429
13.3 Variable Ranking and Statistical Learning 431
13.3.1 Variable Ranking with the Canonical Correlation Matrix C 432
13.3.2 Prediction with a Selected Number of Principal Components 434
13.3.3 Variable Ranking for Discriminant Analysis Based on C 438
13.3.4 Properties of the Ranking Vectors of the Naive $\widehat{C}$ when d Grows 442
13.4 Sparse Principal Component Analysis 449
13.4.1 The Lasso, SCoTLASS Directions and Sparse Principal Components 449
13.4.2 Elastic Nets and Sparse Principal Components 453
13.4.3 Rank One Approximations and Sparse Principal Components 458
13.5 (In)Consistency of Principal Components as the Dimension Grows 461
13.5.1 (In)Consistency for Single-Component Models 461
13.5.2 Behaviour of the Sample Eigenvalues, Eigenvectors and Principal Component Scores 465
13.5.3 Towards a General Asymptotic Framework for Principal Component Analysis 471

Problems for Part III 476

Bibliography 483
Author Index 493
Subject Index 498
Data Index 503