
Contents

| | |
|--|--------------|
| List of Figures | xiii |
| List of Tables | xvii |
| Preface | xxiii |
| 1 Introduction | 1 |
| 1.1 Classification | 1 |
| 1.2 Finite Mixture Models | 3 |
| 1.3 Model-Based Clustering, Classification, and Discriminant Analysis | 4 |
| 1.4 Comparing Partitions | 6 |
| 1.5 R Packages | 8 |
| 1.6 Datasets | 9 |
| 1.7 Outline of the Contents of This Monograph | 9 |
| 2 Mixtures of Multivariate Gaussian Distributions | 11 |
| 2.1 Historical Development | 11 |
| 2.2 Parameter Estimation | 14 |
| 2.2.1 Model-Based Clustering | 14 |
| 2.2.2 Model-Based Classification | 15 |
| 2.2.3 Model-Based Discriminant Analysis | 17 |
| 2.2.4 Initialization via Deterministic Annealing | 18 |
| 2.2.5 Stopping Rules | 18 |
| 2.3 Gaussian Parsimonious Clustering Models | 20 |
| 2.4 Model Selection | 24 |
| 2.5 Merging Gaussian Components | 25 |
| 2.6 Illustrations | 27 |
| 2.6.1 x2 Data | 27 |
| 2.6.2 Banknote Data | 29 |
| 2.6.3 Female Voles Data | 31 |
| 2.6.4 Italian Olive Oil Data | 33 |
| 2.7 Comments | 36 |

| | | |
|----------|---|-----------|
| 3 | Mixtures of Factor Analyzers and Extensions | 39 |
| 3.1 | Factor Analysis | 39 |
| 3.1.1 | The Model | 39 |
| 3.1.2 | An EM Algorithm for the Factor Analysis Model . . . | 40 |
| 3.1.3 | Woodbury Identity | 42 |
| 3.1.4 | Comments | 43 |
| 3.2 | Mixture of Factor Analyzers | 43 |
| 3.3 | Parsimonious Gaussian Mixture Models | 44 |
| 3.3.1 | A Family of Eight Models | 44 |
| 3.3.2 | Parameter Estimation | 44 |
| 3.3.3 | Comments | 48 |
| 3.4 | Expanded Parsimonious Gaussian Mixture Models | 49 |
| 3.4.1 | A Family of Twelve Models | 49 |
| 3.4.2 | Parameter Estimation | 50 |
| 3.5 | Mixture of Common Factor Analyzers | 52 |
| 3.5.1 | The Model | 52 |
| 3.5.2 | Parameter Estimation | 52 |
| 3.5.3 | Discussion | 55 |
| 3.6 | Illustrations | 55 |
| 3.6.1 | x2 Data | 55 |
| 3.6.2 | Italian Wine Data | 56 |
| 3.6.3 | Italian Olive Oil Data | 58 |
| 3.6.4 | Alon Colon Cancer Data | 59 |
| 3.7 | Comments | 61 |
| 4 | Dimension Reduction and High-Dimensional Data | 63 |
| 4.1 | Implicit and Explicit Approaches | 63 |
| 4.2 | PGMM Family in High-Dimensional Applications | 64 |
| 4.3 | VSCC | 65 |
| 4.4 | clustvarsel and selvarclust | 67 |
| 4.5 | GMMDR | 68 |
| 4.6 | HD-GMM | 69 |
| 4.7 | Illustrations | 71 |
| 4.7.1 | Coffee Data | 71 |
| 4.7.2 | Leptograpsus Crabs | 71 |
| 4.7.3 | Banknote Data | 74 |
| 4.7.4 | Wisconsin Breast Cancer Data | 74 |
| 4.7.5 | Leukaemia Data | 75 |
| 4.8 | Comments | 76 |

| | | |
|----------|--|------------|
| 5 | Mixtures of Distributions with Varying Tail Weight | 79 |
| 5.1 | Mixtures of Multivariate t -Distributions | 79 |
| 5.2 | Mixtures of Power Exponential Distributions | 82 |
| 5.3 | Illustrations | 89 |
| 5.3.1 | Overview | 89 |
| 5.3.2 | x2 Data | 89 |
| 5.3.3 | Body Data | 89 |
| 5.3.4 | Diabetes Data | 90 |
| 5.3.5 | Female Voles Data | 92 |
| 5.3.6 | Leptograpsus Crabs Data | 93 |
| 5.4 | Comments | 94 |
| 6 | Mixtures of Generalized Hyperbolic Distributions | 99 |
| 6.1 | Overview | 99 |
| 6.2 | Generalized Inverse Gaussian Distribution | 99 |
| 6.2.1 | A Parameterization | 99 |
| 6.2.2 | An Alternative Parameterization | 100 |
| 6.3 | Mixtures of Shifted Asymmetric Laplace Distributions | 101 |
| 6.3.1 | Shifted Asymmetric Laplace Distribution | 101 |
| 6.3.2 | Parameter Estimation | 102 |
| 6.3.3 | SAL Mixtures versus Gaussian Mixtures | 104 |
| 6.4 | Mixture of Generalized Hyperbolic Distributions | 106 |
| 6.4.1 | Generalized Hyperbolic Distribution | 106 |
| 6.4.2 | Parameter Estimation | 108 |
| 6.5 | Mixture of Generalized Hyperbolic Factor Analyzers | 111 |
| 6.5.1 | The Model | 111 |
| 6.5.2 | Parameter Estimation | 111 |
| 6.5.3 | Analogy with the Gaussian Solution | 114 |
| 6.6 | Illustrations | 115 |
| 6.6.1 | Old Faithful Data | 115 |
| 6.6.2 | Yeast Data | 116 |
| 6.6.3 | Italian Wine Data | 118 |
| 6.6.4 | Liver Data | 118 |
| 6.7 | A Note on Normal Variance-Mean Mixtures | 119 |
| 6.8 | Comments | 120 |
| 7 | Mixtures of Multiple Scaled Distributions | 123 |
| 7.1 | Overview | 123 |
| 7.2 | Mixture of Multiple Scaled t -Distributions | 124 |
| 7.3 | Mixture of Multiple Scaled SAL Distributions | 126 |
| 7.4 | Mixture of Multiple Scaled Generalized Hyperbolic Distribu- tions | 127 |

| | | |
|----------|---|------------|
| 7.5 | Mixture of Coalesced Generalized Hyperbolic Distributions . . . | 128 |
| 7.6 | Cluster Convexity | 129 |
| 7.7 | Illustrations | 132 |
| 7.7.1 | Bankruptcy Data | 132 |
| 7.7.2 | Other Clustering Examples | 133 |
| 7.7.3 | Classification and Discriminant Analysis Examples . . . | 135 |
| 7.8 | Comments | 137 |
| 8 | Methods for Longitudinal Data | 139 |
| 8.1 | Modified Cholesky Decomposition | 139 |
| 8.2 | Gaussian Mixture Modelling of Longitudinal Data | 140 |
| 8.2.1 | The Model | 140 |
| 8.2.2 | Model Fitting | 141 |
| 8.2.2.1 | VEA Model | 141 |
| 8.2.2.2 | EVI Model | 144 |
| 8.2.3 | Constraining Sub-Diagonals of \mathbf{T}_g | 145 |
| 8.2.3.1 | V_d EA Model | 145 |
| 8.2.3.2 | E_d VI Model | 147 |
| 8.2.4 | Modelling the Component Means | 147 |
| 8.3 | Using t -Mixtures | 148 |
| 8.4 | Illustrations | 150 |
| 8.4.1 | Clustering | 150 |
| 8.4.2 | Classification | 153 |
| 8.5 | Comments | 156 |
| 9 | Miscellanea | 157 |
| 9.1 | On the Definition of a Cluster | 157 |
| 9.2 | What Is the Best Way to Perform Clustering, Classification, and Discriminant Analysis? | 159 |
| 9.3 | Mixture Model Averaging | 162 |
| 9.4 | Robust Clustering | 164 |
| 9.5 | Clustering Categorical Data | 166 |
| 9.6 | Cluster-Weighted Models | 168 |
| 9.7 | Mixed-Type Data | 169 |
| 9.7.1 | A Mixture of Latent Variables Model | 169 |
| 9.7.2 | Illustration: Urinary System Disease Diagnosis | 170 |
| 9.8 | Alternatives to the EM Algorithm | 172 |
| 9.8.1 | Variational Bayes Approximations | 172 |
| 9.8.2 | Other Approaches | 173 |
| 9.9 | Challenges and Open Questions | 174 |

| | |
|--|------------|
| <i>Contents</i> | xi |
| Appendix | 177 |
| A.1 Linear Algebra Results | 177 |
| A.2 Matrix Calculus Results | 178 |
| A.3 Method of Lagrange Multipliers | 179 |
| References | 181 |
| Index | 205 |