Finite Mixture Models

G.J. McLACHLAN and D. PEEL

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

To
Beryl, Jonathan, and Robbie

Samantha

Contents

Preface xix

1 General Introduction 1
1.1 Introduction 1
1.1.1 Flexible Method of Modeling 1
1.1.2 Initial Approach to Mixture Analysis 2
1.1.3 Impact of EM Algorithm 3
1.2 Overview of Book 4
1.3 Basic Definition 6
1.4 Interpretation of Mixture Models 7
1.5 Shapes of Some Univariate Normal Mixtures 9
1.5.1 Mixtures of Two Normal Homoscedastic Components 9
1.5.2 Mixtures of Univariate Normal Heteroscedastic Components 11
1.6 Modeling of Asymmetrical Data 14
1.7 Normal Scale Mixture Model 17
1.8 Spurious Clusters 17
1.9 Incomplete-Data Structure of Mixture Problem 19
1.10 Sampling Designs for Classified Data 21
1.11 Parametric Formulation of Mixture Model 22

1.12 Nonparametric ML Estimation of a Mixing Distribution 23
1.13 Estimation of Mixture Distributions 24
1.14 Identifiability of Mixture Distributions 26
1.15 Clustering of Data via Mixture Models 29
1.15.1 Mixture Likelihood Approach to Clustering 29
1.15.2 Decision-Theoretic Approach 30
1.15.3 Clustering of I.I.D. Data 31
1.15.4 Image Segmentation or Restoration 32
1.16 Hidden Markov Models 33
1.17 Testing for the Number of Components in Mixture Models 34
1.18 Brief History of Finite Mixture Models 35
1.19 Notation 37

2 ML Fitting of Mixture Models 40
2.1 Introduction 40
2.2 ML Estimation 40
2.3 Information Matrices 41
2.4 Asymptotic Covariance Matrix of MLE 42
2.5 Properties of MLEs for Mixture Models 42
2.6 Choice of Root 44
2.7 Test for a Consistent Root 44
2.7.1 Basis of Test 44
2.7.2 Example 2.1: Likelihood Function with Two Maximizers 45
2.7.3 Formulation of Test Statistic 45
2.8 Application of EM Algorithm for Mixture Models 47
2.8.1 Direct Approach 47
2.8.2 Formulation as an Incomplete-Data Problem 48
2.8.3 E-Step 48
2.8.4 M-Step 49
2.8.5 Assessing the Implied Error Rates 50
2.9 Fitting Mixtures of Mixtures 51
2.10 Maximum ***a Posteriori*** Estimation 52
2.11 An Aitken Acceleration-Based Stopping Criterion 52
2.12 Starting Values for EM Algorithm 54
2.12.1 Specification of an Initial Parameter Value 54
2.12.2 Random Starting Values 55

2.12.3 Example 2.2: Synthetic Data Set 1 56
2.12.4 Deterministic Annealing EM Algorithm 57
2.13 Stochastic EM Algorithm 61
2.14 Rate of Convergence of the EM Algorithm 61
2.14.1 Rate Matrix for Linear Convergence 61
2.14.2 Rate Matrix in Terms of Information Matrices 62
2.15 Information Matrix for Mixture Models 63
2.15.1 Direct Evaluation of Observed Information Matrix 63
2.15.2 Extraction of Observed Information Matrix in Terms of the Complete-Data Log Likelihood 64
2.15.3 Approximations to Observed Information Matrix: I.I.D. Case 64
2.15.4 Supplemented EM Algorithm 66
2.15.5 Conditional Bootstrap Approach 67
2.16 Provision of Standard Errors 68
2.16.1 Information-Based Methods 68
2.16.2 Bootstrap Approach to Standard Error Approximation 68
2.17 Speeding up Convergence 70
2.17.1 Introduction 70
2.17.2 Louis' Method 71
2.17.3 Quasi-Newton Methods 72
2.17.4 Hybrid Methods 72
2.18 Outlier Detection from a Mixture 74
2.18.1 Introduction 74
2.18.2 Modified Likelihood Ratio Test 74
2.19 Partial Classification 75
2.20 Partial Nonrandom Classification 76
2.20.1 Introduction 76
2.20.2 A Nonrandom Model 77
2.20.3 Asymptotic Relative Efficiencies 77
2.21 Classification ML Approach 79

3 Multivariate Normal Mixtures 81
3.1 Introduction 81
3.2 Heteroscedastic Components 81
3.3 Homoscedastic Components 83
3.4 Standard Errors 83

3.5 Assessment of Model Fit 84
3.6 Examples of Univariate Normal Mixtures 85
3.6.1 Basic Model in Genetics 85
3.6.2 Example 3.1: PTC Sensitivity Data 86
3.6.3 Example 3.2: Screening for Hemochromatosis 87
3.6.4 Example 3.3: Diagnostic Criteria for Diabetes 89
3.7 Examples of Multivariate Normal Mixtures 90
3.7.1 Example 3.4: Crab Data 90
3.7.2 Example 3.5: Hemophilia Data 92
3.8 Properties of MLE for Normal Components 94
3.8.1 Heteroscedastic Components 94
3.8.2 Homoscedastic Components 96
3.9 Options 97
3.9.1 Choice of Local Maximizer 97
3.9.2 Choice of Model for Component-Covariance Matrices 97
3.9.3 Starting Points 98
3.10 Spurious Local Maximizers 99
3.10.1 Introduction 99
3.10.2 Example 3.6: Synthetic Data Set 2 100
3.10.3 Example 3.7: Synthetic Data Set 3 102
3.10.4 Example 3.8: Modeling Hemophilia Data under Heteroscedasticity 103
3.10.5 Detection of Spurious Local Maximizers 103
3.10.6 Example 3.9: Galaxy Data Set 104
3.11 Example 3.10: Prevalence of Local Maximizers 105
3.12 Alternative Models for Component-Covariance Matrices 109
3.12.1 Spectral Representation 109
3.12.2 Example 3.11: Minefield Data Set 110
3.13 Some Other Models 112
3.13.1 Clustering of Treatment Means in ANOVA 112
3.13.2 Three-Way Models 114
3.13.3 Example 3.12: Consumer Data on Cat Food 114
3.13.4 Errors-In-Variables Model 116

4 Bayesian Approach to Mixture Analysis 117
4.1 Introduction 117
4.2 Estimation for Proper Priors 119

4.3 Conjugate Priors 119
4.4 Markov Chain Monte Carlo 120
4.4.1 Posterior Simulation 120
4.4.2 Perfect Sampling 121
4.5 Exponential Family Components 121
4.6 Normal Components 122
4.6.1 Conjugate Priors 122
4.6.2 Gibbs Sampler 123
4.7 Prior on Number of Components 124
4.8 Noninformative Settings 125
4.8.1 Improper Priors 125
4.8.2 Data-Dependent Priors 126
4.8.3 Markov Prior on Component Means 126
4.8.4 Reparameterization for Univariate Normal Components 127
4.9 Label-Switching Problem 129
4.10 Prior Feedback Approach to ML Estimation 132
4.11 Variational Approach to Bayesian Estimation 132
4.12 Minimum Message Length 133

5 Mixtures with Nonnormal Components 135
5.1 Introduction 135
5.2 Mixed Continuous and Categorical Variables 136
5.2.1 Location Model-Based Approach 137
5.2.2 Implementation of Location Model 138
5.3 Example 5.1: Prostate Cancer Data 139
5.3.1 Description of Data Set 139
5.3.2 Fitting Strategy under MULTIMIX 140
5.4 Generalized Linear Model 142
5.4.1 Definition 142
5.4.2 ML Estimation for a Single GLM Component 143
5.4.3 Quasi-Likelihood Approach 144
5.5 Mixtures of GLMs 145
5.5.1 Specification of Mixture Model 145
5.5.2 ML Estimation via the EM Algorithm 146
5.5.3 M-Step 147
5.5.4 Multicycle ECM Algorithm 148
5.5.5 Choice of the Number of Components 148
5.6 A General ML Analysis of Overdispersion in a GLM 149

5.7 Poisson Regression Model 150
5.7.1 Some Standard Modifications for Overdispersed Data 150
5.7.2 Gamma-Poisson Mixture Model 151
5.7.3 Multiplicative Random Effects Model 153
5.7.4 Additive Random Effects Model 153
5.8 Finite Mixture of Poisson Regression Models 154
5.8.1 Mean and Variance 154
5.8.2 Identifiability 155
5.8.3 Example 5.2: Fabric Faults Data Set 155
5.8.4 Components and Mixing Proportions Without Covariates 157
5.8.5 Algorithms for NPMLE of a Mixing Distribution 158
5.8.6 Disease Mapping 158
5.9 Count Data with Excess Zeros 159
5.9.1 History of Problem 160
5.9.2 Zero-Inflated Poisson Regression 160
5.10 Logistic Regression Model 160
5.11 Finite Mixtures of Logistic Regressions 162
5.11.1 Mean and Variance 162
5.11.2 Mixing at the Binary Level 163
5.11.3 Identifiability 164
5.11.4 Example 5.3: Beta-Blockers Data Set 165
5.12 Latent Class Models 166
5.13 Hierarchical Mixtures-of-Experts Model 167
5.13.1 Mixtures-of-Experts Model 167
5.13.2 Hierarchical Mixtures-of-Experts 169
5.13.3 Application of EM Algorithm to HME Model 171
5.13.4 Example 5.4: Speech Recognition Problem 172
5.13.5 Pruning HME Tree Structures 174

6 Assessing the Number of Components in Mixture Models 175
6.1 Introduction 175
6.1.1 Some Practical Issues 175
6.1.2 Order of a Mixture Model 176
6.1.3 Example 6.1: Adjusting for Effect of Skewness on the LRT 177
6.2 Example 6.2: 1872 Hidalgo Stamp Issue of Mexico 179

6.3 Approaches for Assessing Mixture Order 184
6.3.1 Main Approaches 184
6.3.2 Nonparametric Methods 184
6.3.3 Method of Moments 185
6.4 Likelihood Ratio Test Statistic 185
6.4.1 Introduction 185
6.4.2 Example 6.3: Breakdown in Regularity Conditions 186
6.5 Distributional Results for the LRTS 187
6.5.1 Some Theoretical Results 187
6.5.2 Some Simulation Results 189
6.5.3 Mixtures of Two Unrestricted Normal Components 190
6.5.4 Mixtures of Two Exponentials 191
6.6 Bootstrapping the LRTS 192
6.6.1 Implementation 192
6.6.2 Application to Three Real Data Sets 194
6.6.3 Applications in Astronomy 196
6.7 Effect of Estimates on P-Values of Bootstrapped LRTS 198
6.7.1 Some Simulation Results 198
6.7.2 Double Bootstrapping 200
6.8 Information Criteria in Model Selection 202
6.8.1 Bias Correction of the Log Likelihood 202
6.8.2 Akaike's Information Criterion 203
6.8.3 Bootstrap-Based Information Criterion 203
6.8.4 Cross-Validation-Based Information Criterion 205
6.8.5 Minimum Information Ratio Criterion 206
6.8.6 Informational Complexity Criterion 207
6.9 Bayesian-Based Information Criteria 207
6.9.1 Bayesian Approach 207
6.9.2 Laplace's Method of Approximation 208
6.9.3 Bayesian Information Criterion 209
6.9.4 Laplace–Metropolis Criterion 210
6.9.5 Laplace–Empirical Criterion 211
6.9.6 Reversible Jump Method 212
6.9.7 MML Principle 212
6.10 Classification-Based Information Criteria 212
6.10.1 Classification Likelihood Criterion 212
6.10.2 Normalized Entropy Criterion 214

6.10.3 Integrated Classification Likelihood Criterion 215
6.11 An Empirical Comparison of Some Criteria 217
6.11.1 Simulated Set 1 218
6.11.2 Simulated Set 2 218
6.11.3 Simulated Set 3 219
6.11.4 Conclusions from Simulations 220

7 Multivariate t Mixtures 221
7.1 Introduction 221
7.2 Previous Work 222
7.3 Robust Clustering 222
7.4 Multivariate t Distribution 223
7.5 ML Estimation of Mixture of t Distributions 224
7.5.1 Application of EM Algorithm 224
7.5.2 E-Step 225
7.5.3 M-Step 227
7.5.4 Application of ECM Algorithm 229
7.6 Previous Work on M-Estimation of Mixture Components 230
7.7 Example 7.1: Simulated Noisy Data Set 231
7.8 Example 7.2: Crab Data Set 234
7.9 Example 7.3: Old Faithful Geyser Data Set 236

8 Mixtures of Factor Analyzers 238
8.1 Introduction 238
8.2 Principal Component Analysis 239
8.3 Single-Factor Analysis Model 240
8.4 EM Algorithm for a Single-Factor Analyzer 241
8.5 Data Visualization in Latent Space 243
8.6 Mixtures of Factor Analyzers 244
8.7 AECM Algorithm for Fitting Mixtures of Factor Analyzers 245
8.7.1 AECM Framework 245
8.7.2 First Cycle 245
8.7.3 Second Cycle 246
8.7.4 Representation of Original Data 248
8.8 Link of Factor Analysis with Probabilistic PCA 248
8.9 Mixtures of Probabilistic PCAs 250
8.10 Initialization of AECM Algorithm 250

8.11 Example 8.1: Simulated Data 252
8.12 Example 8.2: Wine Data 254

9 Fitting Mixture Models to Binned Data 257
9.1 Introduction 257
9.2 Binned and Truncated Data 258
9.3 Application of EM Algorithm 259
9.3.1 Missing Data 259
9.3.2 E-Step 260
9.3.3 M-Step 261
9.3.4 M-Step for Normal Components 261
9.4 Practical Implementation of EM Algorithm 262
9.4.1 Computational Issues 262
9.4.2 Numerical Integration at Each EM Iteration 262
9.4.3 Integration over Truncated Regions 263
9.4.4 EM Algorithm for Binned Multivariate Data 264
9.5 Simulations 264
9.6 Example 9.1: Red Blood Cell Data 265

10 Mixture Models for Failure-Time Data 268
10.1 Introduction 268
10.2 Competing Risks 269
10.2.1 Mixtures of Survival Functions 269
10.2.2 Latent Failure-Time Approach 270
10.2.3 ML Estimation for Mixtures of Survival Functions 271
10.3 Example 10.1: Heart-Valve Data 272
10.3.1 Description of Problem 272
10.3.2 Mixture Models with Unconstrained Components 273
10.3.3 Constrained Mixture Models 274
10.3.4 Conditional Probability of a Reoperation 276
10.3.5 Advantages of Mixture Model-Based Approach 276
10.4 Long-Term Survivor Model 277
10.4.1 Definition 277
10.4.2 Modified Long-Term Survivor Model 278
10.4.3 Partial ML Approach for Modified Long-Term Survival Model 279

10.4.4 Interpretation of Cure Rate in Presence of Competing Risks 280
10.4.5 Example 9.2: Breast Cancer Data 280
10.5 Analysis of Masked System-Life Data 283
10.5.1 Masked Cause of Failure 283
10.5.2 Application of EM Algorithm 283
10.5.3 Exponential Components 284
10.5.4 Weibull Components 285

11 Mixture Analysis of Directional Data 287
11.1 Introduction 287
11.2 Joint Sets 287
11.3 Directional Data 291
11.4 Initial Work on Clustering of Directional Data 292
11.5 Mixture of Kent Distributions 292
11.6 Moment Estimation of Kent Distribution 293
11.7 Uniform Component for Background Noise 295
11.8 Application of EM Algorithm 296
11.9 Example 11.1: Two Mining Samples 297
11.10 Determining the Number of Joint Sets 298
11.11 Discussion 301

12 Variants of the EM Algorithm for Large Databases 302
12.1 Introduction 302
12.2 Incremental EM Algorithm 303
12.2.1 Introduction 303
12.2.2 Definition of Partial E-Step 303
12.2.3 Block Updating of Sufficient Statistics 303
12.2.4 Justification of IEM Algorithm 305
12.2.5 Gain in Convergence Time 305
12.2.6 IEM Algorithm for Singleton Blocks 306
12.2.7 Efficient Updating Formulas 306
12.3 Simulations for IEM Algorithm 307
12.3.1 Simulation 1 307
12.3.2 Simulation 2 309
12.4 Lazy EM Algorithm 310
12.5 Sparse EM Algorithm 311
12.6 Sparse IEM Algorithm 312
12.6.1 Some Simulation Results 312

12.6.2 Summary of Results for the IEM and SPIEM Algorithms 315
12.7 A Scalable EM Algorithm 316
12.7.1 Introduction 316
12.7.2 Primary Compression of the Data 316
12.7.3 Updating of Parameter Estimates 318
12.7.4 Merging of Sufficient Statistics 319
12.7.5 Secondary Data Compression 319
12.7.6 Tuning Constants 320
12.7.7 Simulation Results 321
12.8 Multiresolution KD-Trees 323
12.8.1 Introduction 323
12.8.2 EM Algorithm Based on Multiresolution KD-Trees 323

13 Hidden Markov Models 326
13.1 Introduction 326
13.2 Hidden Markov Chain 328
13.2.1 Definition 328
13.2.2 Some Examples 329
13.3 Applying EM Algorithm to Hidden Markov Chain Model 329
13.3.1 EM Framework 329
13.3.2 E-Step 330
13.3.3 Forward–Backward Recursions on E-Step 330
13.3.4 M-Step 332
13.3.5 Numerical Instabilities 332
13.4 Hidden Markov Random Field 332
13.4.1 Specification of Markov Random Field 333
13.4.2 Application of EM Algorithm 333
13.4.3 Restoration Step 334
13.4.4 An Improved Approximation to EM Solution 335
13.4.5 Approximate M-Step for Normal Components 336
13.5 Example 13.1: Segmentation of MR Images 336
13.6 Bayesian Approach 338
13.7 Examples of Gibbs Sampling with Hidden Markov Chains 339

Appendix Mixture Software 343

A.1 EMMIX 343
A.2 Some Other Mixture Software 345

References 349

Author Index 395

Subject Index 407

Preface

The importance of finite mixture models in the statistical analysis of data is underscored by the ever-increasing rate at which articles on mixture applications appear in the statistical and general scientific literature. The aim of this monograph is to provide an up-to-date account of the theory and applications of modeling via finite mixture distributions. Since the appearance of the monograph of McLachlan and Basford (1988) on finite mixtures, the literature has expanded enormously to the extent that another monograph on the topic is apt. In the past decade the extent and the potential of the applications of finite mixture models have widened considerably. Because of their flexibility, mixture models are being increasingly exploited as a convenient, semiparametric way in which to model unknown distributional shapes. This is in addition to their obvious applications where there is group-structure in the data or where the aim is to explore the data for such structure, as in a cluster analysis.

In this book the more recent work is surveyed against the background of the existing literature. The widespread use of mixture models in recent times is demonstrated by the fact that of the 800 or so references in this book, almost 40% of them have been published since 1995. A comprehensive account of the major issues involved with modeling via finite mixture distributions is provided. They include identifiability problems, the actual fitting of finite mixtures through use of the EM algorithm, the properties of the maximum likelihood estimators so obtained, the assessment of the number of components to be used in the mixture, and the applicability of asymptotic theory in providing a basis for the solutions to some of these problems. The intent is to provide guidelines to users of mixture models on these various issues. The emphasis is on the applications of mixture models, not only in mainstream statistical analyses, but also in other areas such as unsupervised pattern recognition, speech recognition, and medical imaging.

With the advent of inexpensive, high-speed computers and the simultaneous rapid development in posterior simulation techniques such as Markov chain Monte Carlo

(MCMC) methods for enabling Bayesian estimation to be undertaken, practitioners are increasingly turning to Bayesian methods for the analysis of complicated statistical models. In this book, we consider the latest developments in Bayesian estimation of mixture models.

New topics that are covered in this book include the scaling of the EM algorithm to allow mixture models to be used in data mining applications involving massively huge databases. In the same spirit, there is also an account of the use of the sparse/incremental EM algorithm and of multiresolution kd-trees for speeding up the implementation of the standard EM algorithm for the fitting of mixture models. Another topic concerns the use of hierarchical mixtures-of-experts models as a powerful new approach to nonlinear regression that is a serious competitor to well-known statistical procedures such as MARS and CART. Other recent developments covered include the use of mixture models for handling overdispersion in generalized linear models and proposals for dealing with mixed continuous and categorical variables. In other recent work, there is the proposal to use t components in the mixture model to provide a robust approach to mixture modeling. A further topic is the use of mixtures of factor analyzers which provide a way of fitting mixture models to high-dimensional data. As mixture models provide a convenient basis for the modeling of dependent data by allowing the component-indicator variables to have a Markovian structure, there is also coverage of the latest developments in hidden Markov models, including the Bayesian approach to this problem. Another problem considered is the fitting of mixture models to multivariate data in binned form, which arises in some important medical applications in practice.

The book also covers the latest developments on existing issues with mixture modeling, such as assessing the number of components to be used in a mixture model and the associated problem of determining how many clusters there are in clustering applications with mixture models.

In presenting these latest results, the authors have attempted to draw together the statistical literature with the machine learning and pattern recognition literature.

It is intended that the book should appeal to both applied and theoretical statisticians, as well as to investigators working in the many diverse areas in which relevant use can be made of finite mixture models. It will be assumed that the reader has a fair mathematical or statistical background. The main parts of the book describing the formulation of the finite mixture approach, detailing its methodology, discussing aspects of its implementation, and illustrating its application in many simple statistical contexts should be comprehensible to graduates with statistics as their major subject. The emphasis is on the practical applications of mixture models; and to this end, numerous examples are given.

Chapter 1 begins with a discussion of mixture models and their applications and gives a brief overview of the current state of the area. It also includes a brief history of mixture models.

Chapter 2 focuses on the maximum likelihood fitting of mixture models via the EM algorithm. It covers issues such as the choice of starting values, stopping criteria, the calculation of the observed information matrix, and the provision of standard errors either by information-based methods or by the bootstrap.

Chapter 3 specializes the results in Chapter 2 to mixtures of normal components. Given the tractability of the multivariate normal distribution, it is not surprising that mixture modeling of continuous data is invariably undertaken by normal mixtures. This chapter also discusses the occurrence and identification of so-called spurious local maximizers of the likelihood function, which is an issue with the fitting of normal mixture models with no restrictions on the component-covariance matrices. A number of illustrative examples are presented.

In Chapter 4, we consider the Bayesian approach to the fitting of mixture models. Estimation in a Bayesian framework is now feasible using posterior simulation via recently developed MCMC methods. Bayes estimators for mixture models are well-defined so long as the prior distributions are proper. One main hindrance is that improper priors yield improper posterior distributions. We discuss the use of "partially proper priors," which do not require subjective input for the component parameters, yet the posterior is proper. We also discuss ways of handling other hindrances, including the effect of label switching, which arises when there there is no real prior information that allows one to discriminate between the components of a mixture model belonging to the same parametric family.

In Chapter 5, we consider the fitting of mixture models with nonnormal component densities, including components suitable for mixed feature variables, where some are continuous and some are categorical. The maximum likelihood fitting of commonly used discrete components such as the binomial and Poisson are undertaken within the wider framework of a mixture of generalized linear models (GLMs). The latter also has the capacity to handle the regression case, where the response is allowed to depend on the value of a vector of covariates. The use of mixtures of GLMs for handling overdispersion in a single GLM component is discussed. In work related to mixtures of GLMs, the mixtures-of-experts model is considered, along with its extension, the hierarchical mixtures-of-experts model. This approach which combines aspects of finite mixture models and GLMs provides a comparatively fast learning and good generalization for nonlinear regression problems, including classification.

Chapter 6 is devoted to the estimation of the order of a mixture model. It covers the two main approaches. One way is based on a penalized form of the log likelihood whereby the likelihood is penalized by the subtraction of a term that "penalizes" the model for the number of parameters in it. The other main way for deciding on the order of a mixture model is to carry out a hypothesis test, using the likelihood ratio as the test statistic.

In Chapter 7, we consider the fitting of mixtures of (multivariate) t distributions, as proposed in McLachlan and Peel (1998a) and Peel and McLachlan (2000). The t distribution provides a longer-tailed alternative to the normal distribution. Hence it provides a more robust approach to the fitting of normal mixture models, as observations that are atypical of a component are given reduced weight in the calculation of its parameters. Also, the use of t components gives less extreme estimates of the posterior probabilities of component membership of the mixture model.

In Chapter 8, we consider mixtures of factor analyzers from the perspective of both (a) a method for model-based density estimation from high-dimensional data, and hence for the clustering of such data, and (b) a method for local dimensionality reduction. We also discuss the close link of mixtures of factor analyzers with mixtures of probabilistic principal component analyzers. The mixtures of factor analyzers model enables a normal mixture model to be fitted to high-dimensional data. The number of free parameters is controlled through the dimension q of the latent factor space. It allows thus an interpolation in model complexities from isotropic to full covariance structures without any restrictions.

In Chapter 9, we consider the fitting of finite mixture models to binned and truncated multivariate data by maximum likelihood via the EM algorithm. The solution for an arbitrary number of dimensions of the feature vector is specialized to the case of bivariate normal mixtures.

Chapter 10 is on the use of mixture distributions to model failure-time data in a variety of situations, which occur in reliability and survival analyses. The focus is on the use of mixture distributions to model time to failure in the case of competing risks or failures.

In Chapter 11, a case study is provided to illustrate the use of mixture models in the analysis of multivariate directional data. Mixtures of Kent distributions are used as an aid in joint set identification.

In Chapter 12, we consider methods for improving the speed of the EM algorithm for the maximum likelihood fitting of mixture models to large databases that preserve the simplicity of implementation of the EM in its standard form. They include the incremental version of the EM algorithm, where only a partial E-step is performed before each M-step, and a sparse version, where not all the posterior probabilities of component membership are updated on each iteration. The use of multiresolution kd-trees to speed up the implementation of the E-step is also described. In addition, we consider how the EM algorithm can be scaled to handle very large databases with a limited memory buffer.

In Chapter 13, recent advances on hidden Markov models are covered. Hidden Markov models are increasingly being adopted in applications, since they provide a convenient way of formulating an extension of a mixture model to allow for dependent data. We discuss hidden Markov chain models in the one-dimensional case and hidden Markov random fields in two or higher dimensions.

A brief account of some of the available software for the fitting of mixture models is provided in the Appendix. This account includes a description of the program EMMIX (McLachlan et al., 1999).

The authors wish to thank Dr. Angus Ng for many helpful and insightful discussions on mixture models, and Katrina Monico for her constructive comments and suggestions on drafts of the manuscript. They would also like to acknowledge gratefully financial support from the Australian Research Council. Thanks are due too to the authors and owners of copyrighted material for permission to reproduce tables and figures.

Brisbane, Australia

Geoffrey J. McLachlan
David Peel