Journal of Statistical Software

January 2007, Volume 18, Issue 6. http://www.jstatsoft.org/

Model-based Methods of Classification: Using the mclust Software in Chemometrics

Chris Fraley
University of Washington

Adrian E. Raftery
University of Washington

Abstract

Due to recent advances in methods and software for model-based clustering, and to the interpretability of the results, clustering procedures based on probability models are increasingly preferred over heuristic methods. The clustering process estimates a model for the data that allows for overlapping clusters, producing a probabilistic clustering that quantifies the uncertainty of observations belonging to components of the mixture. The resulting clustering model can also be used for some other important problems in multivariate analysis, including density estimation and discriminant analysis. Examples of the use of model-based clustering and classification techniques in chemometric studies include multivariate image analysis, magnetic resonance imaging, microarray image segmentation, statistical process control, and food authenticity. We review model-based clustering and related methods for density estimation and discriminant analysis, and show how the R package **mclust** can be applied in each instance.

Keywords: model-based clustering, classification, density estimation, discriminant analysis, R, **mclust**.

1. Introduction

Clustering and classification methods are among the most important techniques in multivariate analysis. Due to recent advances in methods and software for model-based clustering, and to the interpretability of the results, clustering procedures based on probability models are increasingly preferred over heuristic methods. Finite mixture models (McLachlan and Peel 2000) provide a principled statistical approach to clustering. Each component probability corresponds to a cluster, and models that differ in the number of components and/or component distributions can be compared using statistical criteria. The clustering process estimates a model for the data that allows for overlapping clusters, producing a probabilistic clustering that quantifies the uncertainty of observations belonging to components of the mixture. The

resulting clustering model can also be used for some other important problems in multivariate analysis, including density estimation and discriminant analysis (Fraley and Raftery 2002). Chemometric studies involving model-based clustering and classification techniques include multivariate image analysis (Wehrens, Simonetti, and Buydens 2002; Wehrens, Buydens, Fraley, and Raftery 2004; Fraley, Raftery, and Wehrens 2005; Tran, Wehrens, and Buydens 2006), magnetic resonance imaging (Wehrens *et al.* 2002; Fraley *et al.* 2005; Forbes, Peyrard, Fraley, Georgian-Smith, Goldhaber, and Raftery 2006), microarray image segmentation (Li, Fraley, Bumgarner, Yeung, and Raftery 2005; Fraley and Raftery 2006c), statistical process control (Thissen, Swierenga, de Weijer, Wehrens, Melssen, and Buydens 2005), and food authenticity (Toher, Downey, and Murphy 2005; Dean, Murphy, and Downey 2006).

In this article, we illustrate model-based clustering, density estimation and discriminant analysis using the R (R Development Core Team 2006) package **mclust** (Fraley and Raftery 1999, 2003, 2006a) available as a contributed package from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/`. We use standard benchmark data sets since our focus is on methodology rather than on specific applications. The first example concerns the diagnosis of diabetes from blood plasma glucose and insulin levels measured under various conditions. The second example involves modeling the atmospheric concentration of carbon monoxide. The third example is the identification of wine cultivars through chemical analysis.

2. Model-based clustering

In model-based clustering, the data x are viewed as coming from a mixture density $f(x) = \sum_{k=1}^{G} \tau_k f_k(x)$, where f_k is the probability density function of the observations in group k, and τ_k is the probability that an observation comes from the kth mixture component ($\tau_k \in (0, 1)$ and $\sum_k^G \tau_k = 1$).

Each component is usually modeled by the normal or Gaussian distribution. Component distributions are characterized by the mean μ_k and the covariance matrix Σ_k, and have the probability density function

$$\phi(x_i; \mu_k, \Sigma_k) = \frac{\exp\{-\frac{1}{2}(x_i - \mu_k)^\top \Sigma_k^{-1}(x_i - \mu_k)\}}{\sqrt{\det(2\pi\Sigma_k)}}.$$

For univariate data, the covariance matrix reduces to a scalar variance. The likelihood for data consisting of n observations assuming a Gaussian mixture model with G multivariate mixture components is

$$\prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \phi(x_i; \mu_k, \Sigma_k). \tag{1}$$

For a fixed number of components G, the model parameters τ_k, μ_k, and Σ_k can be estimated using the EM algorithm initialized by hierarchical model-based clustering (Dasgupta and Raftery 1998; Fraley and Raftery 1998). Data generated by mixtures of multivariate normal densities are characterized by groups or clusters centered at the means μ_k, with increased density for points nearer the mean. The corresponding surfaces of constant density are ellipsoidal. Geometric features (shape, volume, orientation) of the clusters are determined by the covariances Σ_k, which may also be parametrized to impose constraints across components. There are a number of possible parameterizations of Σ_k, many of which are implemented in

the R package **mclust**. Common instances include $\Sigma_k = \lambda I$, where all components are spherical and of the same size; $\Sigma_k = \Sigma$ constant across components, where all components have the same geometry but need not be spherical; and unrestricted Σ_k, where each component may have a different geometry.

Banfield and Raftery (1993) proposed a general framework for geometric constraints in multivariate normal mixtures by parametrizing covariance matrices through eigenvalue decomposition in the following form:

$$\Sigma_k = \lambda_k D_k A_k D_k^\top, \tag{2}$$

where D_k is the orthogonal matrix of eigenvectors, A_k is a diagonal matrix whose elements are proportional to the eigenvalues, and λ_k is an associated constant of proportionality. The decomposition factors λ_k, A_k and D_k are treated as independent sets of parameters, and either constrained to be the same for each component or allowed to vary among components. When parameters are fixed, components share certain geometric properties: D_k governs the orientation of the kth component of the mixture, A_k its shape, and λ_k its volume, which is proportional to $\lambda_k^d \det(A_k)$.

The model options available in the R package **mclust** are summarized in Table 1. In one dimension, there are just two models: `E` for equal variance and `V` for varying variance. In more than one dimension, the model identifiers encode geometric characteristics of the model. For example, `EVI` denotes a model in which the volumes of all clusters are equal (`E`), the shapes of the clusters may vary (`V`), and the orientation is the identity (`I`). Clusters in this model have diagonal covariances with orientation parallel to the coordinate axes. Parameters associated with characteristics designated by `E` or `V` are determined from the data.

A 'best' model can be estimated by fitting models with differing parameterizations and/or

Identifier	Model	# Covariance parameters	Distribution
`EII`	λI	1	Spherical
`VII`	$\lambda_k I$	G	Spherical
`EEI`	λA	d	Diagonal
`VEI`	$\lambda_k A$	$G + (d-1)$	Diagonal
`EVI`	λA_k	$1 + G(d-1)$	Diagonal
`VVI`	$\lambda_k A_k$	Gd	Diagonal
`EEE`	$\lambda D A D^\top$	$d(d+1)/2$	Ellipsoidal
`EEV`	$\lambda D_k A D_k^\top$	$1 + (d-1) + G[d(d-1)/2]$	Ellipsoidal
`VEV`	$\lambda_k D_k A D_k^\top$	$G + (d-1) + G[d(d-1)/2]$	Ellipsoidal
`VVV`	$\lambda_k D_k A_k D_k^\top$	$G[d(d+1)/2]$	Ellipsoidal

Table 1: Parameterizations of the multivariate Gaussian mixture model available in **mclust**. Model identifiers use three letters to encode code geometric characteristics: volume, shape, and orientation. `E` means *equal* and `V` means *varying* across components or clusters; `I` refers to the identity matrix in specifying shape or orientation and is a special case of `E`. In the column labeled '# Covariance parameters', d denotes the dimension of the data, and G denotes the number of mixture components. The total number of parameters for each model can be obtained by adding Gd parameters for the means and $G-1$ parameters for the mixing proportions.

numbers of components to the data by maximum likelihood, and then applying a statistical criterion for model selection. The Bayesian Information Criterion or BIC (Schwarz 1978) is the model selection criterion provided in the **mclust** software. It adds a penalty term on the number of parameters to the loglikelihood. For details of model-based clustering, see McLachlan and Peel (2000) and Fraley and Raftery (2002).

We illustrate model-based clustering on the `diabetes` dataset (Reaven and Miller 1979) giving three measurements for each of 145 subjects described in Table 2.

This dataset is a standard introductory example for model-based clustering (e.g., Banfield and Raftery 1993; Fraley and Raftery 1998, 2006b) and is included in the **mclust** package. The subjects were clinically diagnosed into three groups: normal, chemically diabetic, and overtly diabetic. The diagnosis is given in the first column of the `diabetes` dataset, which is excluded from the cluster analysis.

The following code computes the model using the function **mclust** and plots its BIC: (see Figure 1, upper left):

```
R> library("mclust")
R> data("diabetes")
R> diabetesModel <- Mclust(diabetes[,-1])
R> plot(diabetesModel, diabetes[,-1], what = "BIC")
```

Function `coordProj` can be used to plot the data and **mclust** classification, marking the means and drawing ellipses (with axes) corresponding to the variance for each group (see Figure 1, upper right).

```
R> coordProj(diabetes[,-1], dimens = c(2,3), what = "classification",
             classification = diabetesModel$classification,
             parameters = diabetesModel$parameters)
```

For this data, model-based clustering chooses a model with three components, each having a different covariance. The corresponding three-group classification matches the three clinically diagnosed groups with 88% accuracy.

The uncertainty of a classification can be assessed in model-based clustering and function `coordProj` can be used to display the relative uncertainty of a classification:

```
R> coordProj(diabetes[,-1], dimens = c(2,3), what = "uncertainty",
             uncertainty = diabetesModel$uncertainty,
             parameters = diabetesModel$parameters)
```

Variable	Description
`glucose`	plasma glucose response to oral glucose
`insulin`	plasma insulin response to oral glucose
`sspg`	steady-state plasma glucose (measures insulin resistance)

Table 2: Description of the three measurements given in the diabetes data set (Reaven and Miller 1979) for 145 subjects.

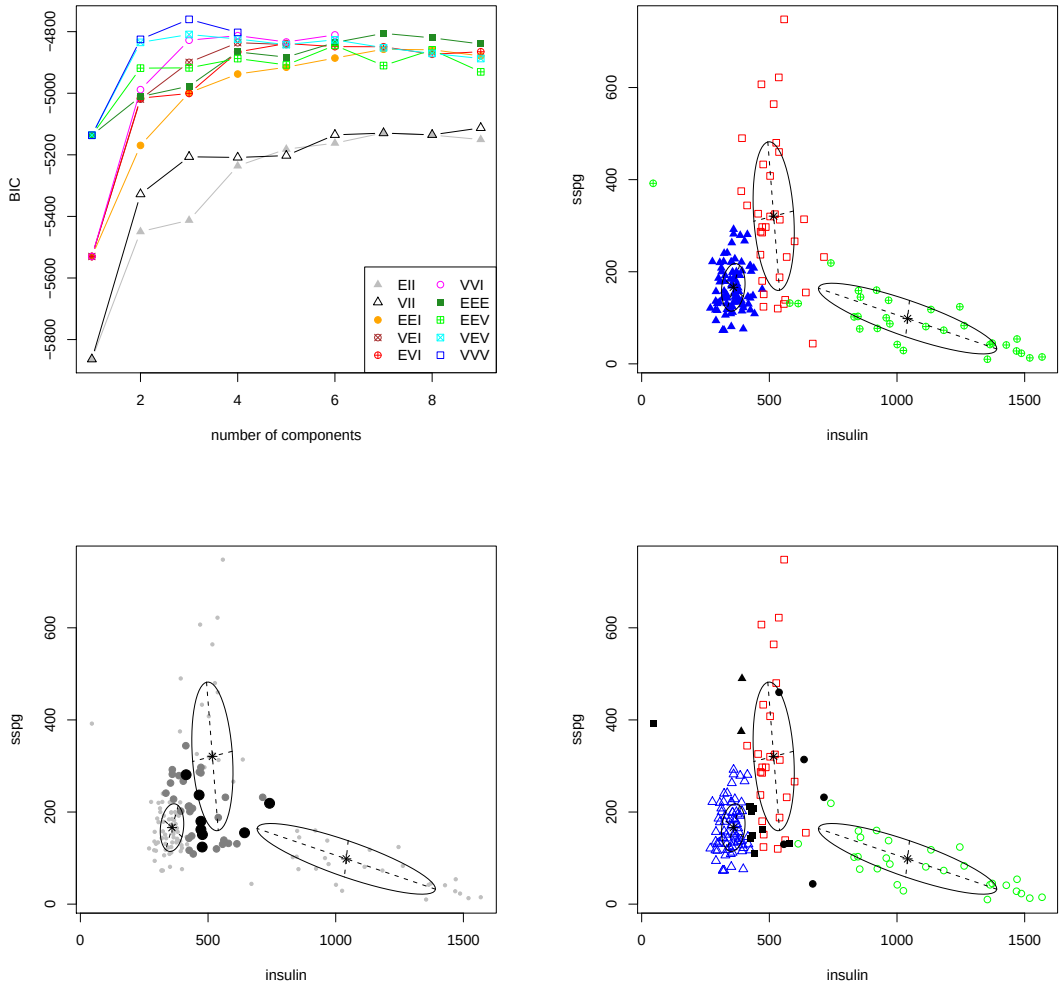

Figure 1: Upper left: BIC from **mclust** for the 10 available model parameterizations and up to 9 clusters for the `diabetes` dataset. Different symbols and line types encode different model parameterizations. The 'best' model is taken to be the one with the highest BIC among the fitted models. Upper right: A projection of the `diabetes` data, with different symbols indicating the classification corresponding to the best model as determined by **mclust**. The component means are marked and ellipses with axes are drawn corresponding to their covariances. In this case there are three components, each with a different covariance. Lower left: A projection of the `diabetes` data showing classification uncertainty. Larger symbols indicate the more uncertain observations. Lower right: A projection of the `diabetes` data showing errors in the **mclust** classification. Filled black symbols indicate incorrectly classificated observations.

The resulting plot is shown in Figure 1, lower left. In this case, the misclassified data points tend to be among the most uncertain, indicating suitability of the cluster model for the data as clinically classified. Since there is a known classification, the classification errors can also be plotted:

```
R> coordProj(diabetes[,-1], dimens = c(2,3), what = "errors",
             classification = diabetesModel$classification,
             parameters = diabetesModel$parameters,
             truth = diabetes[,1])
```

The resulting plot is shown in Figure 1, lower right. Note that some of observations with low uncertainty are misclassified; these to correspond to outliers of the component distributions.

3. Density estimation

While membership in components is important in clustering, the mixture likelihood (1) itself, or its value at given points, is the focus of interest in density estimation (Silverman 1986; Scott 1992). The fitted likelihood can be used for example to reveal or compare data trends. We use the `co2` dataset (Keeling and Whorf 1999) included in the R language to illustrate an application of density estimation. It is a time series of measurements of atmospheric concentrations of carbon dioxide.

Although we treat `co2` as one-dimensional data in this example, we convert it to a matrix to make it easier to extract the observations we need through indexing by months and years:

```
R> co2mat <- matrix(co2, ncol = 12)
R> dimnames(co2mat) <- list(1959:1997, month.abb)
```

We divide the data into three thirteen year periods, and compare the corresponding density estimates.

```
R> earlyData <- as.vector(co2mat[as.character(1959:1971),])
R> middleData <- as.vector(co2mat[as.character(1972:1984),])
R> lateData <- as.vector(co2mat[as.character(1985:1997),])

R> library("mclust")
R> earlyModel <- Mclust(earlyData)
R> middleModel <- Mclust(middleData)
R> lateModel <- Mclust(lateData)

R> mclust1Dplot(earlyData, parameters = earlyModel$parameters,
                what = "density", xlab = "PPM")
R> title("1959 - 1971")
R> mclust1Dplot(middleData, parameters = middleModel$parameters,
                what = "density", xlab = "PPM")
R> title("1972 - 1984")
R> mclust1Dplot(lateData, parameters = lateModel$parameters,
                what = "density", xlab = "PPM")
R> title("1985 - 1997")
```

These plots, shown in Figure 2, indicate that the overall density for the data remains relatively unchanged over the years.

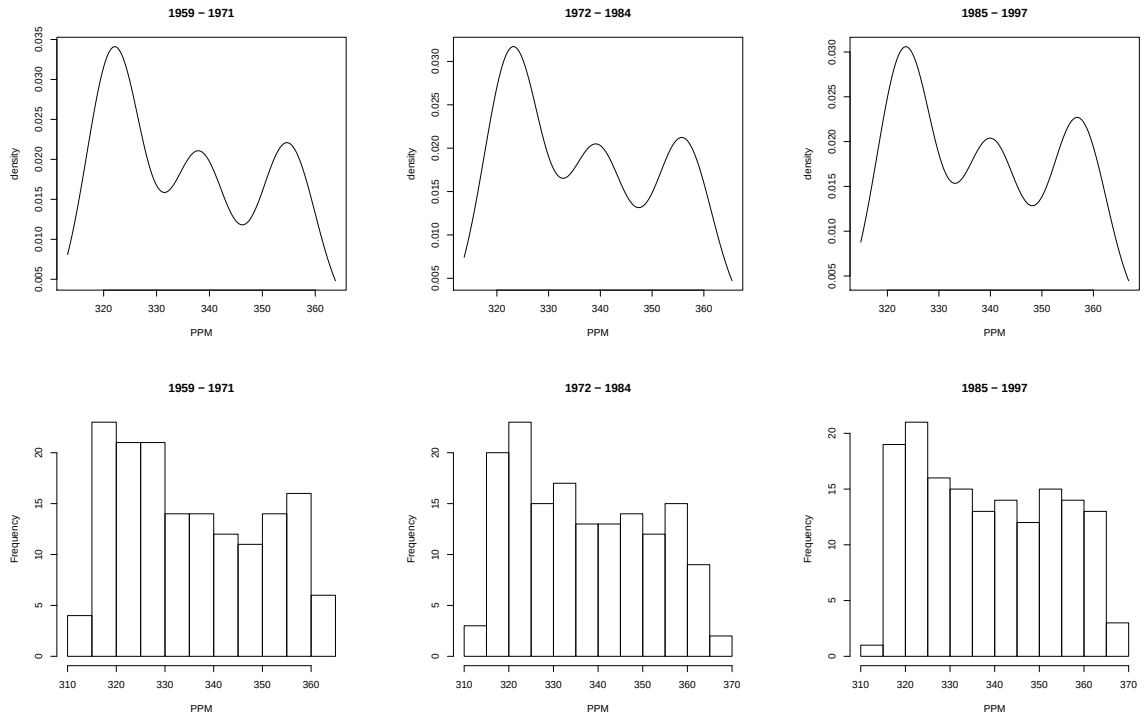

Figure 2: Density via mixture modeling of atmospheric concentrations of carbon dioxide for three 13-year time periods 1959-1971 (left), 1972-1984 (middle), and 1985-1997 (right) and the corresponding histograms. The densities capture the data trends better than the histograms, and show that the distribution of concentrations is essentially the same for each time period.

More generally, density estimates at specific values can be obtained via function `dens`. As an illustration, we divide the data into cold and warm months, and compare the corresponding density estimates. The following computes and plots density estimates for the model fit to the data in the warm and cold months:

```
R> coldData <- as.vector(co2mat[,c("Oct","Nov","Dec","Jan","Feb","Mar")])
R> warmData <- as.vector(co2mat[,c("Apr","May","Jun","Jul","Aug","Sep")])

R> library("mclust")
R> coldModel <- Mclust(coldData)
R> warmModel <- Mclust(warmData)

R> coldDens <- dens(modelName=coldModel$modelName, data = ppm,
                    parameters = coldModel$parameters )
R> warmDens <- dens(modelName=warmModel$modelName, data = ppm,
                    parameters = warmModel$parameters )

R> ylim <- range(c(coldDens, warmDens))
R> x <- seq(from = min(ppm), to = max(ppm))
```

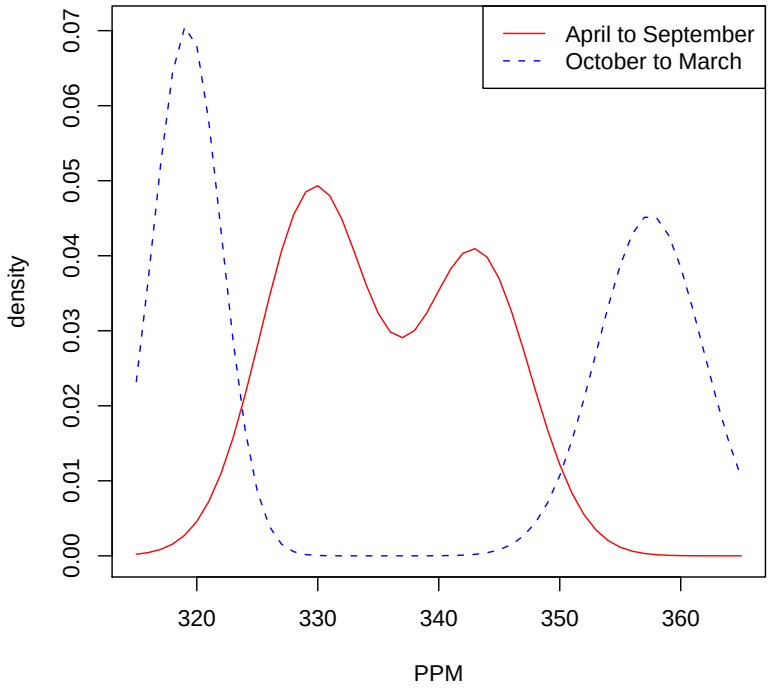

Figure 3: Density via mixture modeling of atmospheric concentrations of carbon dioxide for warm (April - September) and cold (October - March) months during the years 1959-1997.

```
R> plot(x, warmDens, ylim=ylim, type="l", ylab="density", xlab="PPM",
        col="red", lty=1)
R> lines(x, coldDens, ylim=ylim, type="l", ylab="density", xlab="PPM",
         col="blue", lty=2)
R> legend("topright", legend = c("April to September","October to March"),
           lty = c(1,2), col = c("red", "blue"))
```

The resulting plots are shown in Figure 3. It is easy to see, for example, that an observations between 330 and 340 ppm would be more likely to occur in the warmer months than in the colder ones, while an observations below 320 ppm and above 350 ppm would be more likely to occur in the colder months.

When comparing density estimates, the question naturally arises as to whether a density estimate can be used to classify data.

4. Discriminant analysis

In *discriminant analysis*, also know as *supervised classification*, known classfications of some observations (the 'training set') are used to classify others. Many methods have been proposed for discriminant analysis, and it is applicable in a wide variety of settings (see e.g., Ripley 1996;

Duda, Hart, and Storck 2001; Hastie, Tibshirani, and Friedman 2001; McLachlan, Bean, and Peel 2002). Discriminant analysis methods are often probabilistic, based on the assumption that observations in each class are generated by a distribution specific to that class. If K is the number of classes, $f_k(\cdot)$ is the probability distribution of the kth class, and τ_k is the proportion of members of the population that are in class k, then according to Bayes' theorem, the posterior probability that an observation $\mathbf{x}$ belongs to the kth class is

$$Pr(\mathbf{x} \in \text{class } k) = \frac{\tau_k f_k(\mathbf{x})}{\sum_{l=1}^{K} \tau_l f_l(\mathbf{x})}.$$

In probabilistic discriminant analysis, a model is fit to each class in the training set, and data is assigned to the class corresponding to the model in which it has the highest posterior probability. When the model is a normal mixture fitted by model-based clustering, the procedure is known as `mclustDA` (Fraley and Raftery 2002).

We illustrate `mclustDA` on the wine recognition database (Forina, Lanteri, Armanino, and Leardi 1998) from the UCI Machine Learning Repository (Newman, Hettich, Blake, and Merz 1998). These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. Since the classification (according to cultivar) is available as the first column of the dataset, but there are no designated training or test sets, we use the odd numbered observations as a training set to classify the even numbered observations.

```
R> odd <- seq(from = 1, to = nrow(wine), by = 2)
R> even <- seq(from = 2, to = nrow(wine), by = 2)

R> wineDA1 <- mclustDA(train = list(data = wine[odd,-1],
                                    labels = wine[odd,1]),
                       test = list(data = wine[even,-1],
                                   labels = wine[even,1]))

R> wineDA1

Modeling Summary:
  trainClass mclustModel numGroups
1          1         VEI         2
2          2         EEI         2
3          3         VEI         2

Test Classification Summary:
 1  2  3
25 40 24

Training Classification Summary:
 1  2  3
30 35 24
```

```
Training Error: 0

Test Error: 0.04494382
```

In this case all of the training models have diagonal covariance matrices (oriented along the 13 coordinate axes), and there is no training error. The test error is less than half a percent. Note that labels need not be provided for the test data: they do not affect the test classification results, but they do allow assessment of the test error if they are available.

The above analysis implicity assumed equal prior probabilities for each class in training data. We could, for example, assume that the prior probabilities are equal to the proportions of the observations in each class in the training data:

```
R> tab <- table(wine[odd,1])
R> tab
 1  2  3
30 35 24

R> pro <- tab/sum(tab)
R> pro
        1         2         3
0.3370787 0.3932584 0.2696629

R> wineDA2 <- mclustDA(train = list(data = wine[odd,-1],
                                    labels = wine[odd,1]),
                       test = list(data = wine[even,-1],
                                   labels = wine[even,1]),
                       pro = pro)
```

In this case the classification results are the same as they were when the probabilities were assumed equal, although this will not necessarily be the case in general.

5. Summary and future work

We have demonstrated the use of the **mclust** software for model-based clustering, density estimation, and discriminant analysis on benchmark datasets involving chemometric measurements. We gave references in the introduction to model-based clustering applications in chemometrics. The current version of **mclust** also includes a Bayesian prior for regularization for datasets in which maximum likelihood estimation fails due to singularities or degeneracies (Fraley and Raftery 2005). There has been considerable recent work on extension to large data sets (Wehrens *et al.* 2004; Fraley *et al.* 2005), to high-dimensional data (McLachlan, Peel, and Bean 2003; Raftery and Dean 2006), and to models involving categorical data (Handcock, Raftery, and Tantrum 2005; Handcock, Tantrum, Shortreed, and Hoff 2006). Model-based clustering is an active area of research, and advances in these and other areas can be anticipated in the near future.

References

Banfield JD, Raftery AE (1993). "Model-based Gaussian and Non-Gaussian Clustering." *Biometrics*, **49**, 803–821.

Dasgupta A, Raftery AE (1998). "Detecting Features in Spatial Point Processes with Clutter via Model-based Clustering." *Journal of the American Statistical Association*, **93**, 294–302.

Dean N, Murphy TB, Downey G (2006). "Updating Classification Rules with Unlabeled Data with Applications in Food Authenticity Studies." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **55**, 1–14.

Duda RO, Hart PE, Storck D (2001). *Pattern Classification*. Wiley, 2nd edition.

Forbes F, Peyrard N, Fraley C, Georgian-Smith D, Goldhaber D, Raftery A (2006). "Model-based Region-of-Interest Selection in Dynamic Breast MRI." *Journal of Computer Assisted Tomography*, **30**(4), 675–687.

Forina M, Lanteri S, Armanino C, Leardi R (1998). "**PARVUS**: An Extendable Package of Programs for Data Exploration, Classification, and Correlation." Institute of Pharmaceutical and Food Analysis Technologies, Genoa, Italy.

Fraley C, Raftery AE (1998). "How Many Clusters? Which Clustering Method? – Answers via Model-based Cluster Analysis." *Computer Journal*, **41**, 578–588.

Fraley C, Raftery AE (1999). "**mclust**: Software for Model-based Cluster Analysis." *Journal of Classification*, **16**, 297–306.

Fraley C, Raftery AE (2002). "Model-based Clustering, Discriminant Analysis and Density Estimation." *Journal of the American Statistical Association*, **97**, 611–631.

Fraley C, Raftery AE (2003). "Enhanced Software for Model-based Clustering, Density Estimation, and Discriminant Analysis: **mclust**." *Journal of Classification*, **20**, 263–286.

Fraley C, Raftery AE (2005). "Bayesian Regularization for Normal Mixture Estimation and Model-based Clustering." *Technical Report 486*, University of Washington, Department of Statistics. (to appear in *Journal of Classification*).

Fraley C, Raftery AE (2006a). "Model-based Microarray Image Analysis." *R News*, **6**, 60–63.

Fraley C, Raftery AE (2006b). "**mclust** Version 3 for R: Normal Mixture Modeling and Model-based Clustering." *Technical Report 504*, University of Washington, Department of Statistics.

Fraley C, Raftery AE (2006c). "Some Applications of Model-based Clustering in Chemistry." *R News*, **6**, 17–23.

Fraley C, Raftery AE, Wehrens R (2005). "Incremental Model-based Clustering for Large Datasets with Small Clusters." *Journal of Computational and Graphical Statistics*, **14**, 1–18.

Handcock MS, Raftery AE, Tantrum J (2005). "Model-based Clustering for Social Networks." *Working Paper no. 46*, Center for Statistics and the Social Sciences.

Handcock MS, Tantrum J, Shortreed S, Hoff P (2006). "**latentnet**: Latent Position and Cluster Models for Statistical Networks." R package version 0.7-15.

Hastie T, Tibshirani R, Friedman JH (2001). *The Elements of Statistical Learning*. Springer-Verlag.

Keeling C, Whorf T (1999). "Atmospheric CO_2 Records from Sites in the SIO Air Sampling Network." In "Trends: A Compendium of Data on Global Change," Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy.

Li Q, Fraley C, Bumgarner RE, Yeung KY, Raftery AE (2005). "Donuts, Scratches and Blanks: Robust Model-based Segmentation of Microarray Images." *Bioinformatics*, **21**, 2875–2882.

McLachlan GJ, Bean RW, Peel D (2002). "A Mixture Model-based Approach to the Clustering of Microarray Gene Expression Data." *Bioinformatics*, **18**, 413–422.

McLachlan GJ, Peel D (2000). *Finite Mixture Models*. Wiley.

McLachlan GJ, Peel D, Bean RW (2003). "Modelling High-Dimensional Data by Mixtures of Factor Analyzers." *Computational Statistics & Data Analysis*, **41**, 379–388.

Newman DJ, Hettich S, Blake CL, Merz C (1998). "UCI Repository of Machine Learning Databases." URL http://www.ics.uci.edu/~mlearn/MLRepository.html.

Raftery AE, Dean N (2006). "Variable Selection for Model-based Clustering." *Journal of the American Statistical Association*, **101**, 168–178.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Reaven GM, Miller RG (1979). "An Attempt to Define the Nature of Chemical Diabetes Using a Multidimensional Analysis." *Diabetologia*, **16**, 17–24.

Ripley BD (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Schwarz G (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, **6**, 461–464.

Scott DW (1992). *Multivariate Density Estimation*. Wiley.

Silverman BW (ed.) (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Thissen U, Swierenga H, de Weijer AP, Wehrens R, Melssen WJ, Buydens LMC (2005). "Multivariate Statistical Process Control Using Mixture Modelling." *Journal of Chemometrics*, **19**, 23–31.

Toher D, Downey G, Murphy T (2005). "A Comparison of Model-based and Regression Classification Techniques Applied to Near-Infrared Spectroscopic Data in Food Authentication Studies." *Technical Report 5/10*, Department of Statistics, Trinity College Dublin.

Tran TN, Wehrens R, Buydens LMC (2006). "**SMIXTURE**: A Strategy of Mixture Models Clustering of Multivariate Images." *Journal of Chemometrics*, **19**, 607–614.

Wehrens R, Buydens L, Fraley C, Raftery A (2004). "Model-based Clustering for Image Segmentation and Large Datasets via Sampling." *Journal of Classification*, **21**, 231–253.

Wehrens R, Simonetti A, Buydens L (2002). "Mixture-Modeling of Medical Magnetic Resonance Data." *Journal of Chemometrics*, **16**, 1–10.

Affiliation:

Chris Fraley and Adrian E. Raftery
Department of Statistics
Box 354322, University of Washington
Seattle, WA 98195–4322, United States of America
E-mail: fraley@stat.washington.edu
URL: http://www.stat.washington.edu/fraley/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 18, Issue 6 *Submitted:* 2006-09-27
January 2007 *Accepted:* 2007-01-10