Home > SIS 2013 Statistical Conference > **Advances in Latent Variables - Methods, Models and Applications**

Advances in Latent Variables - Methods, Models and Applications

Brescia - Department of Economics and Management

June 19, 2013 – June 21, 2013

Electronic Book "Advances in Latent Variables"

The full version of the Conference Proceedings, collected in the **Electronic Book "Advances in Latent Variables"**, Eds Brentari E., Carpita M., Vita e Pensiero, Milan, Italy, ISBN 978 88 343 2556 8, are available here.

SPEC-M1.2 - Latent variable models for marketing research

Market segmentation via mixtures of constrained factor analyzers **PDF**

Francesca Greselin, Salvatore Ingrassia

A Bayesian nonparametric model for data on different scale of measure; an application to customer base management of telecommunications companies PDF

Antonio Canale, David B. Dunson

Separating between- and within- group associations and effects for categorical variables PDF

Marcel August Croon, Margot Bennink

Market segmentation via mixtures of constrained factor analyzers

Francesca Greselin and Salvatore Ingrassia

Abstract In this paper we introduce a procedure for the parameter estimation of mixtures of factor analyzers, which maximizes the likelihood function in a constrained parameter space, to overcome the well known issue of singularities and to reduce spurious maxima of the likelihood function. A Monte Carlo study of the performance of the algorithm is provided. Finally the proposed approach is employed to provide a market segmentation, to model a set of quantitative variables provided by a telecom company, and related to the amount of services used by customers.

Key words: Market segmentation, Mixture of Factor Analyzers, Model-Based Clustering, Constrained EM algorithm.

1 Introduction and motivation

Finite mixture distributions, dating back to the seminal works of Newcomb and Pearson, have been receiving a growing interest in statistical modeling all along the last century. Their central role is mainly due to their double nature: they combine the flexibility of non-parametric models with the strong and useful mathematical properties of parametric models. In such finite mixture models, it is assumed that a sample of observations arises from a specified number of underlying populations of unknown proportions and the purpose is to decompose the sample into its mixture components.

Beyond these ”unconditional” approaches to finite mixtures of normal distributions, ”conditional” mixture models allow for the simultaneous probabilistic classification of observations and the estimation of regression models relating covariates to the expectations of the dependent variable within latent classes. In more detail,

Francesca Greselin
Department of Statistics and Quantitative Methods
Milano-Bicocca University
Via Bicocca degli Arcimboldi 8 - 20126 Milano (Italy), e-mail: francesca.greselin@unimib.it

Salvatore Ingrassia
Department of Economics and Business
University of Catania
Corso Italia 55 - Catania (Italy), e-mail: s.ingrassia@unict.it

these expectations are specified as linear functions of a set of explanatory variables. A large number of mixture regression models has been developed (see Wedel and DeSarbo, 1994, for a review, and also e.g. Wedel and Kamakura (2000), Wedel and DeSarbo (2002)). This methodology is particularly employed in marketing research, mainly due to the availability of categorical and ordinal data originated by surveys.

On the other hand, when quantitative variables are available, the "unconditional" approach has been also extensively considered in the literature. Our paper aims at providing a market segmentation, based on mixtures of gaussian factor analyzers, to model a set of quantitative variables provided by a telecom company, and related to the amount of services used by customers. Along the lines of Ghahramani and Hilton (1997) we assume that the data have been generated by a linear factor model with latent variables modeled as Gaussian mixtures. Our purpose is to improve the performances of the EM algorithm, giving practical recipes to overcome some of its issues. It is well known that the EM algorithm generates a sequence of estimates, starting from an initial guess, so that the corresponding sequence of the log-likelihood values is not decreasing. However, the convergence toward the MLE is not guaranteed, because the log-likelihood is unbounded and presents local maxima. Following Ingrassia (2004), in this paper we introduce and implement a procedure for the parameters estimation of mixtures of factor analyzers, which maximizes the likelihood function in a constrained parameter space, having no singularities and a reduced number of spurious local maxima.

We have organized the rest of the paper as follows. In Section 2 we summarize main ideas about Gaussian Mixtures of Factor Analyzer model; in Section 3 we give basic notes on likelihood function and the AECM algorithm. Some well known considerations (Hathaway, 1985) related to spurious maximizers and singularities in the EM algorithm are recalled in Section 4, and motivate our proposal to introduce constraints on factor analyzers. Further, we give a detailed methodology to implement such constraints. In Section 5 we show and discuss the improved performance of our procedure via simulations, while in Section 6 an application to market Segmentation is presented. Section 7 contains concluding notes.

2 The Gaussian Mixture of Factor analyzers

Within the Gaussian Mixture (GM) model-based approach to density estimation and clustering, the density of the d-dimensional random variable $\mathbf{X}$ of interest is modeled as a mixture of a number, say G, of multivariate normal densities in some unknown proportions $\pi_1, \ldots \pi_G$,

$$f(\mathbf{x};\theta) = \sum_{g=1}^{G} \pi_g \phi_d(\mathbf{x};\mu_g,\Sigma_g) \tag{1}$$

where $\phi_d(\mathbf{x};\mu,\Sigma)$ denotes the d-variate normal density function with mean μ and covariance matrix Σ. Here the vector $\theta_{GM}(d,G)$ of unknown parameters consists of the $(G-1)$ mixing proportions π_g, the $G \times d$ elements of the component means μ_g, and the $\frac{1}{2}Gd(d+1)$ distinct elements of the component-covariance matrices Σ_g.

Then, we postulate a finite mixture of linear sub-models for the distribution of the full observation vector $\mathbf{X}$, given the (unobservable) latent factors $\mathbf{U}$

$$\mathbf{X}_i = \mu_g + \Lambda_g \mathbf{U}_{ig} + \mathbf{e}_{ig} \quad \text{with probability} \quad \pi_g \, (g = 1, \ldots, G) \quad \text{for } i = 1, \ldots, n, \tag{2}$$

where Λ_g is a $d \times q$ matrix of *factor loadings*, the *factors* $\mathbf{U}_{1g}, \ldots, \mathbf{U}_{ng}$ are $\mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ distributed independently of the *errors* $\mathbf{e}_{ig}$, which are independently $\mathcal{N}(\mathbf{0}, \Psi_g)$ distributed, and Ψ_g is a $d \times d$ diagonal matrix $(g = 1, \ldots, G)$. We suppose that $q < d$, which means that q latent factors are jointly explaining the d observable features of the statistical units. Under these assumptions, the mixture of factor analyzers model is given by (1), where the g-th component-covariance matrix Σ_g has the form

$$\Sigma_g = \Lambda_g \Lambda_g' + \Psi_g \quad (g = 1, \ldots, G).$$

The parameter vector $\theta_{MGFA}(d, q, G)$ now consists of the elements of the component means μ_g, the Λ_g, and the Ψ_g, along with the mixing proportions π_g $(g = 1, \ldots, G-1)$. Note that, as $q(q-1)/2$ constraints are needed for Λ_g to be uniquely defined, the number of free parameters, for each component of the mixture, is reduced to $dq + d - \frac{1}{2}q(q-1)$.

3 The likelihood function and the EM algorithm for MGFA

In this section we summarize the main steps of the EM algorithm for mixtures of Factor analyzers, see e.g. McLachlan *et al*. (2003) for details.

Let $\mathbf{x}_i$ $(i = 1, \ldots, n)$ denotes the realization of $\mathbf{X}_i$ in (2). Then, the complete-data likelihood function for a sample $\underset{\sim}{\mathbf{X}}$ of size n can be written as

$$L_c(\theta; \underset{\sim}{\mathbf{X}}) = \prod_{i=1}^{n} \prod_{g=1}^{G} \left[\phi_d \left(\mathbf{x}_i | \mathbf{u}_i; \mu_g, \Lambda_g, \Psi_g \right) \phi_q(\mathbf{u}_{ig}) \pi_g \right]^{z_{ig}}. \tag{3}$$

Due to the factor structure of the model, we consider the alternating expectation-conditional maximization (AECM) algorithm, consisting of the iteration of two conditional maximizations, until convergence. There is one E-step and one CM-step, alternatively i) considering $\theta_1 = \{\pi_g, \mu_g, g = 1, \ldots, G\}$ where the missing data are the unobserved group labels $\underset{\sim}{\mathbf{Z}} = (\mathbf{z}_1', \ldots, \mathbf{z}_n')$ and ii) considering $\theta_2 = \{(\Lambda_g, \Psi_g), g = 1, \ldots, G\}$ where the missing data are the group labels $\mathbf{Z}$ and the unobserved latent factors $\mathbf{U} = (\mathbf{U}_{11}, \ldots, \mathbf{U}_{nG})$.

In the First Cycle, after updating the $z_{ig}^{(k+1)}$ in the E-step, the M-step provides new values for $\pi_g^{(k+1)}, \mu_g^{(k+1)}, n_g^{(k+1)}$.

In the Second Cycle, after writing the complete data log-likelihood, some algebras lead to the following estimate of $\{(\Lambda_g, \Psi_g), g = 1, \ldots, G\}$

$$\hat{\Lambda}_g = \mathbf{S}_g^{(k+1)} \gamma_g^{(k)\prime} [\Theta_g^{(k)}]^{-1} \qquad \hat{\Psi}_g = \operatorname{diag} \left\{ \mathbf{S}_g^{(k+1)} - \hat{\Lambda}_g \gamma_g^{(k)} \mathbf{S}_g^{(k+1)} \right\},$$

where

$$\mathbf{S}_g^{(k+1)} = (1/n_g^{(k+1)}) \sum_{i=1}^{n} z_{ig}^{(k+1)} (\mathbf{x}_i - \mu_g^{(k+1)})(\mathbf{x}_i - \mu_g^{(k+1)})'$$
$$\gamma_g^{(k)} = \Lambda_g^{(k)'} (\Lambda_g^{(k)} \Lambda_g^{(k)'} + \Psi_g^{(k)})^{-1}$$
$$\Theta_{ig}^{(k)} = \mathbf{I}_q - \gamma_g^{(k)} \Lambda_g^{(k)} + \gamma_g^{(k)} (\mathbf{x}_i - \mu_g)(\mathbf{x}_i - \mu_g)' \gamma_g^{(k)'}.$$

Hence the ML estimates $\hat{\Lambda}_g$ and $\hat{\Psi}_g$ for Λ and Ψ, given an initial random clustering $\mathbf{z}^{(0)}$, can be obtained iteratively, carrying out the following steps, for $g = 1, \ldots, G$:

1. Compute $z_{ig}^{(k+1)}$ and consequently obtain $\pi_g^{(k+1)}$, $\mu_g^{(k+1)}$, $n_g^{(k+1)}$ and $\mathbf{S}_g^{(k+1)}$;
2. Set a starting value for Λ_g and Ψ_g from $\mathbf{S}_g^{(k+1)}$;
3. Repeat the following steps, until convergence on $\hat{\Lambda}_g$ and $\hat{\Psi}_g$:
   a. Compute $\gamma_g^+ = \Lambda_g'(\Lambda_g \Lambda_g' + \Psi_g)^{-1}$ and $\Theta_g^+ = \mathbf{I}_q - \gamma_g \Lambda_g + \gamma_g \mathbf{S}_g^{(k+1)} \gamma_g'$;
   b. Set $\gamma_g \leftarrow \gamma_g^+$ and $\Theta_g \leftarrow \Theta_g^+$;
   c. Compute $\Lambda_g^+ \leftarrow \mathbf{S}_g^{(k+1)} \gamma_g' (\Theta_g^{-1})$ and $\Psi_g^+ \leftarrow \operatorname{diag}\left\{\mathbf{S}_g^{(k+1)} - \Lambda_g^+ \gamma_g \mathbf{S}_g^{(k+1)}\right\}$;
   d. Set $\Lambda_g \leftarrow \Lambda_g^+$ and $\Psi_g \leftarrow \Psi_g^+$;

4 Likelihood maximization in constrained parametric spaces

Maximum likelihood estimation for normal mixture models have been shown to be an ill-posed problem, because $\mathscr{L}(\theta)$ is unbounded on Θ. Moreover, the likelihood may present many local maxima and any small number of sample points, grouped sufficiently close together or even lying roughly in a subspace, can give raise to spurious maximizers. To overcome this issues, Hathaway (1985) proposed a constrained maximum likelihood formulation for mixtures of univariate normal distributions, suggesting a natural extension to the multivariate case. Let $c \in (0,1]$, then the following constraints

$$\min_{1 \le h \neq j \le k} \lambda(\Sigma_h \Sigma_j^{-1}) \ge c \tag{4}$$

on the eigenvalues λ of $\Sigma_h \Sigma_j^{-1}$ leads to properly defined, scale-equivariant, consistent ML-estimators for the mixture-of-normal case, see Hennig (2004). It is easy to show that a sufficient condition for (4) is

$$a \le \lambda_{ig} \le b, \qquad i = 1, \ldots, d; \qquad g = 1, \ldots, G \tag{5}$$

where λ_{ig} denotes the ith eigenvalue of Σ_g i.e. $\lambda_{ig} = \lambda_i(\Sigma_g)$, and for $a, b \in \mathbb{R}^+$ such that $a/b \ge c$, see Ingrassia (2004). Differently from (4), condition (5) can be easily implemented in any optimization algorithm.

Applying the eigenvalue decomposition to the square $d \times d$ matrix $\Lambda_g \Lambda_g'$ we can find Γ_g and Δ_g such that $\Lambda_g \Lambda_g' = \Gamma_g \Delta_g \Gamma_g'$ where Γ_g is the orthonormal matrix whose rows are the eigenvectors of $\Lambda_g \Lambda_g'$ and $\Delta_g = \operatorname{diag}(\delta_{1g}, \ldots, \delta_{dg})$ is the diagonal matrix of the eigenvalues of $\Lambda_g \Lambda_g'$, sorted in non increasing order, i.e. $\delta_{1g} \ge \delta_{2g} \ge \ldots \ge \delta_{qg} \ge 0$, and $\delta_{(q+1)g} = \cdots = \delta_{dg} = 0$. Further, applying now the singular value decomposition to Λ_g, we get $\Lambda_g = \mathbf{U}_g\mathbf{D}_g\mathbf{V}_g'$, where $\mathbf{U}_g$ and $\mathbf{V}_g$ are unitary matrices whose columns are respectively the *left* and *right singular vectors* of Λ_g, $\mathbf{D}_g$ is a $d \times q$ rectangular matrix with the q *singular values* $\{d_{1g},\ldots,d_{qg}\}$ of Λ_g on the diagonal. This yields

$$\Lambda_g\Lambda_g' = (\mathbf{U}_g\mathbf{D}_g\mathbf{V}_g')(\mathbf{V}_g\mathbf{D}_g'\mathbf{U}_g') = \mathbf{U}_g\mathbf{D}_g\mathbf{D}_g'\mathbf{U}_g' \tag{6}$$

and we get $\Gamma_g = \mathbf{U}_g$ and $\Delta_g = \mathbf{D}_g\mathbf{D}_g'$, that is $\mathrm{diag}(\delta_{1g},\ldots,\delta_{qg}) = \mathrm{diag}(d_{1g}^2,\ldots,d_{qg}^2)$. We can now modify the EM algorithm in such a way that the eigenvalues of the covariances Σ_g (for $g = 1,\ldots,G$) are confined into suitable ranges. To this aim we exploit the following inequalities

$$\lambda_{\min}(\Lambda_g\Lambda_g' + \Psi_g) \geq \lambda_{\min}(\Lambda_g\Lambda_g') + \lambda_{\min}(\Psi_g) \geq a$$
$$\lambda_{\max}(\Lambda_g\Lambda_g' + \Psi_g) \leq \lambda_{\max}(\Lambda_g\Lambda_g') + \lambda_{\max}(\Psi_g) \leq b$$

which enforce (5) when imposing the following constraints

$$d_{ig}^2 + \psi_{ig} \geq a \qquad i = 1,\ldots,d \tag{7}$$
$$d_{ig} \leq \sqrt{b - \Psi_{ig}} \qquad i = 1,\ldots,q \tag{8}$$
$$\psi_{ig} \leq b \qquad i = q+1,\ldots,d \tag{9}$$

for $g = 1,\ldots,G$, where ψ_{ig} denotes the i-th diagonal entry of Ψ_g. In particular, we note that condition (7) reduces to $\Psi_{ig} \geq a$ for $i = (q+1),\ldots,d$.

It is important to remark that the resulting EM algorithm is monotone, once the initial guess, say Σ_g^0, satisfies the constraints. Further, as shown in the case of gaussian mixtures in Ingrassia and Rocci (2007), the maximization of the complete loglikelihood is guaranteed. From the other side, it is apparent that the above recipes require some a priori information on the covariance structure of the mixture, throughout the bounds a and b.

5 Numerical studies

To show the performance of the constrained EM algorithm we present here a brief numerical study. More simulations, also with real datasets, can be found in Greselin and Ingrassia (2013). A sample of $N = 150$ data has been generated with weights $\pi = (0.3, 0.4, 0.3)'$ according to the following parameters:

$$\mu_1 = (0,0,0,0,0,0)' \qquad \Psi_1 = \mathrm{diag}(0.1,0.1,0.1,0.1,0.1,0.1)$$
$$\mu_2 = (5,5,5,5,5,5)' \qquad \Psi_2 = \mathrm{diag}(0.4,0.4,0.4,0.4,0.4,0.4)$$
$$\mu_3 = (10,10,10,10,10,10)' \qquad \Psi_3 = \mathrm{diag}(0.2,0.2,0.2,0.2,0.2,0.2)$$

$$\Lambda_1 = \begin{pmatrix} 0.50 & 1.00 \\ 1.00 & 0.45 \\ 0.05 & -0.50 \\ -0.60 & 0.50 \\ 0.50 & 0.10 \\ 1.00 & -0.15 \end{pmatrix} \quad \Lambda_2 = \begin{pmatrix} 0.10 & 0.20 \\ 0.20 & 0.50 \\ 1.00 & -1.00 \\ -0.20 & 0.50 \\ 1.00 & 0.70 \\ 1.20 & -0.30 \end{pmatrix} \quad \Lambda_3 = \begin{pmatrix} 0.10 & 0.20 \\ 0.20 & 0.00 \\ 1.00 & 0.00 \\ -0.20 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & -1.30 \end{pmatrix}.$$

We note that $\max_{i,g} \lambda_i(\Sigma_g) = 4.18$. The parameters provided by the EM algorithm, when it is initialized from the true classification, will be considered as the point of local maximum corresponding to the consistent estimator θ^*. In the following, such estimate will be referred to as the right maximum of the likelihood function. We run a hundred times both the unconstrained and the constrained AECM algorithms (for different values of the constraints a, b) using a common random initial clusterings, and the algorithms run until convergence or they reach the fixed maximum number (150) of iterations. The stopping criterion is based on the Aitken acceleration procedure.

The unconstrained algorithm attains the right solution in 24% of cases (see Figure 1).We set the bound $a = 0.01$ to protect from small eigenvalues in the estimated covariance matrices; conversely, as local maxima are quite often due to large estimated eigenvalues, we consider different values for b, the upper bound. To compare how a and b influences the performance of the constrained EM, different pairs of values has been considered, and Table 1 shows the more interesting cases. Further results are reported in Figure 1, where the boxplots of the distribution of the misclassification errors show the poor performance of the unconstrained algorithm compared to its constrained version. For all values of the upper bound b, the third quartile of the misclassification error is steadily equal to 0. Indeed, for $b = 6, 10$ and 15 we had no misclassification error, and we observed very low values for it when $b = 20$ and $b = 25$ (respectively 3 and 11 non null values, over 100 runs). Moreover, the robustness of the results with respect to the choice of the upper constraint is apparent.

Table 1 Percentage of convergence to the right maximum of the constrained EM algorithms for $a = 0.01$ and some values of the upper constraint b

b	$+\infty$	6	10	15	20	25
	24%	100%	100%	100%	97%	89%

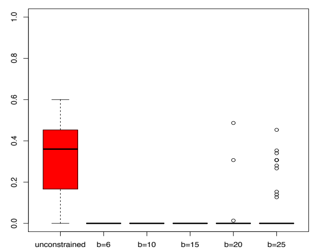

Fig. 1 Boxplots of the misclassification error. From left to right, boxplots refer to the unconstrained algorithm, then to the constrained algorithm, for $a = 0.01$ and $b = 6, 10, 15, 20, 25$.

In Figure 2 we plot the classified data on the three factor spaces given by $\hat{\mathbf{U}}_{i1}, \hat{\mathbf{U}}_{i2}$ and $\hat{\mathbf{U}}_{i3}$ under the true maximum of the likelihood function (first rows of plots), while in the second row we give the classification obtained according to a spurious maximum of the likelihood function: only the appropriate factor space allows for the right classification, providing a strong motivation toward using latent variables.

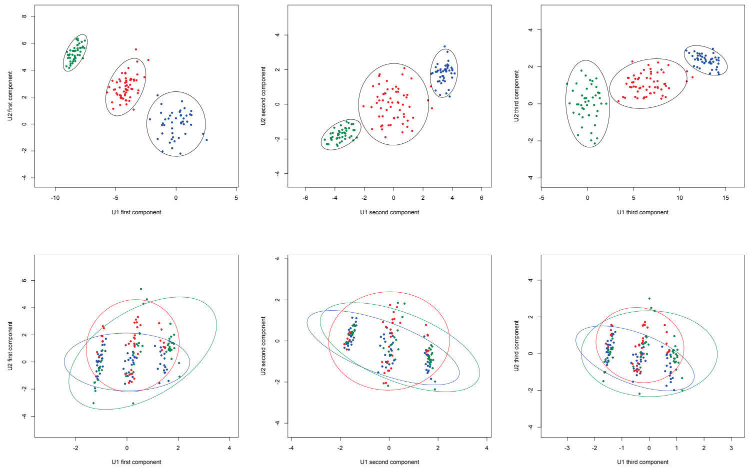

Fig. 2 Plot of the classified data on the three factor spaces, under the true maximum of the likelihood function (upper row) and, conversely, under a spurious maximum (row below)

6 An application to market segmentation

In this section we deal with multivariate data provided by a telecom company, and related to the amount of services used by customers. Our aim is to assess if customers with the same traffic plan have the same behavior or subgroups with different usage can be observed among them. We have a sample of 1449 customers and we deal with quantitative variables about total traffic usage (over 6 mths: Aug12-Jan13), like minutes of voice call, number of events of voice call, number of sent SMS, number of events and amount of data downloaded from Internet. The data are divided into ”traffic below the threshold of the plan” or ”out of it”, ”traffic On” or ”Off net”, and so on, summing up to 45 variables. To select the more important variables for the subsequent analysis, we adopted the random forest methodology, in the classification setting. This pre-step selects 7 final variables (with a loss of about 3.24% in terms of the Out-of-Box estimate of error rate, which increased from 16.55% to 19.79%). They are the downloaded Kilobytes, the number of sent SMS Off and On net, the duration of Voice calls to Fixed, to mobile Off and On net. Our aim is to discover if the bimodal densities we observe in some of the univariate data distributions of the variables is well interpreted by the mixture of gaussian factor model, in such a way that a non-unique underlying behavior of customers can be confirmed to the market analysts. For the choice of the upper bound b, we adopted a data driven method. Let $\mathbf{X}$ denote the matrix of the data, we run firstly the constrained algorithm with $b = +\infty$, then we re-run it with a grid of $k = 10$ values from $\lambda^* = \lambda_{\max}(\mathrm{Cov}(\mathbf{X})) = 11.56587$ till λ^*/k stopping when we get a decrease in the final likelihood. Iterating this approach, say a hundred times, we observe that the algorithm provide a data-driven choice for $b = 6.9395$ (analogously for the lower bound a). Now, we want to compare our results with the Mixture of Common Factor Analyzers (MCFA) (McLachlan *et al.*, 2003), which is the milestone in the literature of mixture gaussian factors. Our estimation provides a mixture of two components as the best solution, with proportions $\pi_1 = 0.754$ and $\pi_2 = 0.246$, reducing from $d = 7$ to $q = 4$ dimensions the data, with BIC$= \ln L - k\ln(n) = -22338.05$. On the other side, MCFA estimates a model with lower BIC$= -25496.44$, i.e. the constraints adopted by the latter model are too strong for the dataset at hand.

7 Concluding remarks

Mixtures of factor analyzers are commonly used when looking for a few latent factors able to describe the underlying structure of the data. In this paper we implemented a methodology to maximize the likelihood function in a constrained parameter space, to overcome some issues of the EM algorithm. The performance of the new estimation approach has been assessed by simulations: results show that the problematic convergence of the EM, even more critical when dealing with factor analyzers, can be greatly improved. Finally, in search of latent variables useful for market segmentation, we provide an application to multivariate data provided by a telecom company.

References

Ghahramani, Z. and Hilton, G. (1997). The EM algorithm for mixture of factor analyzers. *Tech. Rep. CRG-TR-96-1*.

Greselin, F. and Ingrassia, S. (2013) Maximum likelihood estimation in constrained parameter spaces for mixtures of factor analyzers, *http://arxiv.org/abs/1301.1505*.

Hathaway, R. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Ann. Stat.*, **13**(2), 795–800.

Hennig, C. (2004). Breakdown points for maximum likelihood-estimators of location-scale mixtures. *Ann. Stat.*, **32**, 1313–1340.

Ingrassia, S. (2004). A likelihood-based constrained algorithm for multivariate normal mixture models. *Stat. Meth. & Appl.*, **13**, 151–166.

Ingrassia, S. and Rocci, R. (2007). Constrained monotone em algorithms for finite mixture of multivariate gaussians. *Comp. Stat. & Data Anal.*, **51**, 5339–5351.

McLachlan, G., Peel, D., and Bean, R. (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Comp. Stat. and Data Anal.*, **41**, 379–388.

Wedel, M., DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *J. Class.*, **12**(1), 21-55.

Wedel, M., DeSarbo W. (2002). Market segment derivation and profiling via a finite mixture model framework, *Market. Lett.*, **13**, 17–25.

Wedel, M., Kamakura, W.A. (2000). *Market Segmentation. Conceptual and Methodological Foundations*. Kluwer Academic Publishers, Boston.