# Clustering fuzzy data using the fuzzy EM algorithm

Benjamin Quost[1] and Thierry Denœux[1]

laboratoire HeuDiaSyC, Université de Technologie de Compiègne
Centre de Recherches de Royallieu, B.P. 20529
F-60205 Compiègne Cedex
{quostben,tdenoeux}@hds.utc.fr

**Abstract.** In this article, we address the problem of clustering imprecise data using finite mixtures of Gaussians. We propose to estimate the parameters of the mixture model using the fuzzy EM algorithm. This extension of the EM algorithm allows us to handle imprecise data represented by fuzzy numbers. First, we briefly recall the principle of the fuzzy EM algorithm. Then, we provide the update equations for the parameters of a Gaussian mixture model for fuzzy data. Experiments carried out on synthetic and real data demonstrate the interest of our approach for clustering data that are only imprecisely known.

## 1 Introduction

Gaussian mixture modelling is a very powerful tool for estimating a multivariate distribution [19]. This model assumes the data to arise from a random sample, whose distribution is a finite mixture of Gaussians. The major difficulty is to estimate the parameters of the model. Generally, these estimates are computed using the maximum-likelihood (ML) approach, through an iterative procedure known as the EM algorithm. Once the parameter values are known, the posterior probabilities of each data point may be computed. Then, classifying each point into the class with highest posterior probability gives a partition of the data. The choice of Gaussian mixtures, rather than geometrical models, is motivated by several arguments. Additional assumptions, for example regarding the shape or the volume of the classes, may be easily taken into account, giving birth to parsimonious variants of the general model. This approach also provides a theoretical framework in which solutions to complex problems, such that determining the number of classes or validating the structure of the partition obtained, may be proposed.

When estimating the parameters using the EM algorithm, the observed data are assumed to be precisely known. However, in some applications, the precise value taken by the variables may be difficult or even impossible to know. For example, acoustic emission control may be used to detect flaws on pressure equipments. This technique provides locations of acoustic events associated with imprecision degrees [8]. The interest of taking into account uncertainty measurements has been demonstrated [16]. Many works advocate the use of fuzzy sets theory for dealing with imprecise data [12, 13, 15, 23]. Some of them consider that the data at hand are intrinsically fuzzy, a position that has been known as the *physical interpretation* of fuzziness. Here, we rather adopt an *epistemic interpretation*, in which fuzzy numbers "imperfectly specify a value that is existing and precise, but not measurable with exactitude under the given observation conditions" [12]. In this setting, a data sample is a collection of possibility distributions. Each one represents the partial knowledge of the precise value taken by the random variable of interest. The problem of clustering fuzzy data has been addressed in a number of recent papers [2, 6, 10, 11, 20–22, 27, 28]. These approaches differ from the type of data considered and from the clustering approach used. However, to our knowledge, clustering imprecise data using mixtures of distributions has only been addressed in [9], when data are intervals.

In this paper, we propose to fit a Gaussian mixture model to the fuzzy data at hand. The likelihood of the sample may be computed using Zadeh's extension principle [29]; then, an EM-like procedure may be used to estimate the parameters maximizing this likelihood. Denœux [4] recently proposed an extension of the EM algorithm for imprecise data in the framework of belief functions. As a possibility distribution may be identified with the plausibility function of a consonant belief mass, this extension is also valid for fuzzy data [5].

The paper is organized as follows. In Section 2, the Gaussian mixture model for clustering data is briefly recalled, along with the procedure for estimating the parameters using the EM algorithm. We focus on the particular case where the covariance matrices are diagonal. In Section 3, we present the fuzzy EM algorithm for estimating the parameters of a Gaussian mixture model with diagonal covariance matrices, when the data are fuzzy numbers. Section 4 presents the experiments on synthetic and real data, and we conclude in Section 5.

## 2 Gaussian mixtures models for crisp data

Here, we recall the main results of Gaussian mixture modeling using the EM algorithm. More information on Gaussian mixture models may be found in [19, 14]. For a thorough study of the EM algorithm, the reader may refer to [17].

### 2.1 Model

We suppose that $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is the realization of a random sample $(X_1, \ldots, X_n)$. Each $\mathbf{x}_i$ is a $p$-dimensional vector: $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, supposed to be drawn from a mixture of $g$ Gaussians of probability density function (pdf):

$$g(\mathbf{x}; \Psi) = \sum_{k=1}^{g} \pi_k g_k(\mathbf{x}; \Psi_k), \tag{1}$$

where $\Psi \in \Omega$ is the vector of parameters of the model, and $g_k(\mathbf{x}; \Psi_k)$ denotes the $k$th Gaussian component with parameters $\Psi_k = (\mathbf{m}_k, \Sigma_k, \pi_k)^\top$:

$$g_k(\mathbf{x}; \Psi_k) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp\left( -\frac{1}{2} (\mathbf{x} - \mathbf{m}_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mathbf{m}_k) \right). \tag{2}$$

Let the $g$-dimensional vector $\mathbf{z}_i = (z_{i1}, \dots, z_{ig})$ indicate the membership of $\mathbf{x}_i$: $z_{ik} = 1$ if $\mathbf{x}_i$ was generated by the $k^{\text{th}}$ component, and 0 otherwise. Now, let us introduce the notations $\mathbf{y} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, and let $(\mathbf{y}, \mathbf{z})$ be the complete data sample, with pdf $g_c(\mathbf{y}, \mathbf{z}; \Psi)$. The EM algorithm aims at maximizing the observed data log-likelihood $L(\Psi) = \sum_{\mathbf{z}} g(\mathbf{y}|\mathbf{z}; \Psi)\,\mathbb{P}(\mathbf{z}; \Psi)$. The algorithm solves this problem by proceeding iteratively with the complete data log-likelihood $\log L_c$. In the case of Gaussian mixtures, we have:

$$\begin{aligned}\log L_c(\Psi) = \log g_c(\mathbf{y}, \mathbf{z}; \Psi) &= \sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik} \log \pi_k + \sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik} \log g_k(\mathbf{x}_i; \Psi_k), \\ &= \sum_{k=1}^{g} \log \pi_k \sum_{i=1}^{n} z_{ik} - \frac{np}{2}\log(2\pi) - \sum_{k=1}^{g}\sum_{i=1}^{n} \frac{z_{ik}}{2} \log |\Sigma_k| \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik} (\mathbf{x}_i - \mathbf{m}_k)^\top \Sigma_k^{-1} (\mathbf{x}_i - \mathbf{m}_k). \end{aligned} \quad (3)$$

In this article, we restrict ourselves to the particular case where the variables are independent conditionally to each class: by definition,

$$g_k(\mathbf{x}; \Psi_k^{(q)}) = \prod_{j=1}^{p} g_k(x_j; \Psi_k^{(q)}); \quad (4)$$

this is equivalent to requiring that the covariance matrices be diagonal: for each $k = 1, \dots, g$, we have $\Sigma_k = \text{diag}(\sigma_{1k}^2, \dots, \sigma_{pk}^2)$, where $\text{diag}(\mathbf{u})$ denotes the matrix whose diagonal is the vector $\mathbf{u}$. The complete log-likelihood thus becomes:

$$\begin{aligned}\log L_c(\Psi) = \sum_{k=1}^{g} \log \pi_k \sum_{i=1}^{n} z_{ik} - \frac{np}{2}\log(2\pi) - \sum_{k=1}^{g}\sum_{i=1}^{n} z_{ik} \sum_{j=1}^{p} \log(\sigma_{jk}) \\ - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{g}\sum_{j=1}^{p} \frac{z_{ik}}{\sigma_{jk}^2}(x_{ij} - m_{jk})^2. \end{aligned} \quad (5)$$

### 2.2 Estimating the parameters using the EM algorithm

The EM algorithm estimates the parameters so as to maximize the likelihood of the observed data. For this purpose, it proceeds iteratively with the complete log-likelihood $\log L_c(\Psi)$, alternating between two steps that we briefly recall here.

**E-step of the EM algorithm** The E-step consists in computing

$$Q(\Psi, \Psi^{(q)}) = \mathbb{E}_{\Psi^{(q)}}[\log L_c(\Psi)|\mathbf{y}, \mathbf{z}]; \quad (6)$$

here, $\Psi^{(q)}$ denotes the current fit of $\Psi$ at iteration $q$, and $\mathbb{E}_{\Psi^{(q)}}$ represents the expectation computed using parameters $\Psi^{(q)}$. Let

$$t_{ik} = \mathbb{E}_{\Psi^{(q)}}[Z_{ik}|\mathbf{x}_i, \Psi_k] = \mathbb{P}_{\Psi^{(q)}}[Z_{ik} = 1|\mathbf{x}_i] = \frac{\pi_k\, g_k(\mathbf{x}_i; \Psi_k)}{\sum_{k=1}^{g} \pi_k\, g_k(\mathbf{x}_i; \Psi_k)}; \quad (7)$$

then,

$$Q(\Psi, \Psi^{(q)}) = \sum_{k=1}^{g} \log \pi_k \sum_{i=1}^{n} t_{ik} - \frac{np}{2} \log(2\pi) - \sum_{k=1}^{g} \sum_{i=1}^{n} t_{ik} \sum_{j=1}^{p} \log(\sigma_{jk}) - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{g} \sum_{j=1}^{p} \frac{t_{ik}}{\sigma_{jk}^2} (x_{ij} - m_{jk})^2. \quad (8)$$

**M-step of the EM algorithm** The M-step then consists in maximizing the expectation $Q(\Psi, \Psi^{(q)})$ with respect to $\Psi^{(q)}$; that is, in computing $\Psi_{q+1}$ such that $Q(\Psi_{q+1}, \Psi^{(q)}) \geq Q(\Psi, \Psi^{(q)}),$ for all $\Psi \in \Omega$. In practice, the update equations are given by setting the derivatives of $Q(\Psi, \Psi^{(q)})$ with respect to each component of $\Psi$ to zero. Assume that the covariance matrices are diagonal (4); then:

$$\pi_k^{(q+1)} = \frac{1}{n} \sum_{i=1}^{n} t_{ik},$$

$$m_{jk}^{(q+1)} = \frac{\sum_{i=1}^{n} t_{ik} x_{ij}}{\sum_{i=1}^{n} t_{ik}},$$

$$\sigma_{jk}^{(q+1)} = \sqrt{\frac{\sum_{i=1}^{n} t_{ik} \left(x_{ij} - m_{jk}^{(q+1)}\right)^2}{\sum_{i=1}^{n} t_{ik}}}.$$

*Remark 1 (Spherical model).* Assume that in each class, the variances of all the variables are equal: $\Sigma_k = \sigma_k^2 \mathrm{Id_p}$, for $k = 1, \ldots, g$. (Here, $\mathrm{Id_p}$ is the $p \times p$ identity matrix.) In this case, the level curves of the density are hyper-spheres, and the classes are said to be spherical. Then, the update equations for the proportions and the means are unchanged; the standard deviations are updated using:

$$\sigma_k^{(q+1)} = \sqrt{\frac{\sum_{i=1}^{n} t_{ik} \sum_{j=1}^{p} \left(x_{ij} - m_{jk}^{(q+1)}\right)^2}{p \sum_{i=1}^{n} t_{ik}}}. \quad (9)$$

**Convergence of the EM algorithm** The convergence of the EM algorithm to a local maximum for the observed log-likelihood $L$ has been proved in [3, 24]. Under some conditions on the initial values of the parameters, $L$ is bounded from above. As the observed log-likelihood increases at each iteration of the algorithm [3], the convergence is ensured. In practice, the algorithm is stopped when the difference between two successive values of $L(\Psi)$ is less than a given threshold $\epsilon$:

$$\log L(\Psi^{(q+1)}) - \log L(\Psi^{(q)}) \leq \epsilon. \quad (10)$$

As noted in [17, page 85], in many practical applications, the EM algorithm converges to a local maximizer of the observed log-likelihood. However, it is underlined that this convergence towards nontrivial solutions relies on the compactness of the parameter space. This assumption may not hold in certain cases.

For example, when computing ML estimators of the parameters in a mixture of Gaussians, setting the mean of a class to be one of the data points and letting its variance tend to zero will let $L(\Psi)$ tend to infinity.

To avoid such degenerate solutions, prior knowledge on the actual value of the parameters may be integrated in the estimation process, using an adequate distribution $p(\Psi)$. Then, the maximum *a posteriori* (MAP) estimate of the vector parameter $\Psi$ may be computed so as to maximize the log (incomplete) posterior density $\log p(\Psi|\mathbf{y},\mathbf{z}) = \log L_c(\Psi) + \log p(\Psi)$. The analytic formulation for the update equations of the parameter estimates is simpler if $p(\Psi)$ is a *conjugate prior* for the distribution of the model. In the case of a mixture of Gaussians, the conjugate prior for a covariance matrix $\Sigma$ is the inverse-Wishart distribution:

$$f(\Sigma) = \frac{|\Lambda|^{m/2}|\Sigma|^{-(m+p+1)/2}\exp\left(-\operatorname{trace}\left(\Lambda\Sigma^{-1}\right)/2\right)}{2^{mp/2}\Gamma_p(m/2)}, \tag{11}$$

where $\Gamma_p$ stands for the ($p$-dimensional) multivariate Gamma distribution, $m \geq p$ is the number of degrees of freedom, and $\Lambda$ is a positive definite matrix. The mean and the mode of this pdf are $\Lambda/(m-p-1)$ and $\Lambda/(m+p+1)$, respectively.

# 3 Gaussian mixture models for fuzzy data

## 3.1 The fuzzy EM algorithm applied to Gaussian mixtures

Here, we briefly present the fuzzy EM (FEM) algorithm [5] that may be derived from Denœux's EM algorithm for credal data [4]. Assume that the available data are imprecise and represented using fuzzy numbers: instead of a crisp value $\mathbf{x}_i$, we have a sample $\tilde{\mathbf{y}}$ of fuzzy numbers, of which each element $\tilde{\mathbf{x}}_i$ has a membership function $\mu_{\tilde{\mathbf{x}}_i}$. The value $\mu_{\tilde{\mathbf{x}}_i}(\mathbf{x})$ may be interpreted as the degree of possibility that the actual value taken by the random variable $X_i$ is $\mathbf{x}$. Thus, the complete-data sample is now $(\tilde{\mathbf{y}},\mathbf{z})$. Then, Zadeh's definition of the probability of a fuzzy event [30] may be used to compute the observed data log-likelihood:

$$L(\Psi) = \sum_{\mathbf{z}} \mathbb{P}(\mathbf{z};\Psi)\int g(\mathbf{y}|\mathbf{z};\Psi)d\mathbf{y}. \tag{12}$$

Thus, the E-step now consists in computing

$$Q(\Psi,\Psi^{(q)}) = \mathbb{E}_{\Psi^{(q)}}\left[\log L_c(\Psi)|\tilde{\mathbf{y}},\mathbf{z}\right]. \tag{13}$$

Note that the expectation is now taken with respect to the fuzzy sample $\tilde{\mathbf{y}}$. We remind here that the conditional density of a continuous random variable $X$ with pdf $g_X$, with respect to a fuzzy event $\tilde{\mathbf{x}}$ with fuzzy membership function $\mu_{\tilde{\mathbf{x}}}$, is:

$$g_X(\mathbf{x}|\tilde{\mathbf{x}}) = \frac{\mu_{\tilde{\mathbf{x}}}(\mathbf{x})g_X(\mathbf{x})}{\int \mu_{\tilde{\mathbf{x}}}(\mathbf{x})g_X(\mathbf{x})d\mathbf{x}}. \tag{14}$$

The M-step still consists, at iteration $q$, in maximizing $Q(\Psi,\Psi^{(q)})$ with respect to $\Psi$. The FEM algorithm iterates alternatedly between steps E and M, until the difference between two successive values is small. Its convergence has been proved [4, 5], using similar arguments to those proposed in [3].

### 3.2 Update equations of the parameters

We describe here how mixtures of Gaussians may be fit to fuzzy data using the FEM algorithm. In addition to the conditional independence of the variables, we assume that the membership function of a multidimensional fuzzy number may be expressed as the product of the membership functions of its components:

$$\mu_{\tilde{\mathbf{x}}_i}(\mathbf{x}) = \prod_{j=1}^{p} \mu_{\tilde{x}_{ij}}(x_j). \tag{15}$$

**E-step of the FEM algorithm** Let us compute $Q(\Psi, \Psi^{(q)}) = \mathbb{E}[\log L_c(\Psi)|\tilde{\mathbf{y}}, \mathbf{z}]$:

$$\begin{aligned} Q(\Psi, \Psi^{(q)}) = & \sum_{k=1}^{g} \log \pi_k \sum_{i=1}^{n} \mathbb{E}_{\Psi^{(q)}}[Z_{ik}|\tilde{\mathbf{x}}_i] - \frac{1}{2} \sum_{k=1}^{g} \log |\Sigma_k| \sum_{i=1}^{n} \mathbb{E}_{\Psi^{(q)}}[Z_{ik}|\tilde{\mathbf{x}}_i] \\ & - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{g} \mathbb{E}_{\Psi^{(q)}} \left[ \sum_{j=1}^{p} \frac{Z_{ik}}{\sigma_{jk}^2} (x_{ij} - m_{jk})^2 |\tilde{\mathbf{x}}_i \right] - \frac{np}{2} \log(2\pi). \end{aligned} \tag{16}$$

Let us introduce the following notations:

$$\gamma_{ik}^{(q)} = \mathbb{P}_{\Psi^{(q)}}(\tilde{\mathbf{x}}_i | Z_{ik} = 1) = \int \mu_{\tilde{\mathbf{x}}_i}(\mathbf{x}) g_k(\mathbf{x}; \Psi_k^{(q)}) d\mathbf{x}, \tag{17}$$

$$\gamma_{ijk}^{(q)} = \mathbb{P}_{\Psi^{(q)}}(\tilde{x}_{ij} | Z_{ik} = 1) = \int \mu_{\tilde{x}_{ij}}(w_j) g_k(w_j; \Psi_k^{(q)}) dw_j; \tag{18}$$

$$p_i^{(q)} = \mathbb{P}_{\Psi^{(q)}}(\tilde{\mathbf{x}}_i) = \sum_{k=1}^{p} \pi_k^{(q)} \int \mu_{\tilde{\mathbf{x}}_i}(\mathbf{x}) g_k(\mathbf{x}; \Psi_k^{(q)}) d\mathbf{x}; \tag{19}$$

$$\eta_{ijk}^{(q)} = \mathbb{E}_{\Psi^{(q)}}[x_{ij} | \tilde{x}_{ij}, Z_{ik} = 1] = \frac{\displaystyle\int x_j \mu_{\tilde{x}_{ij}}(x_j) g_{jk}(x_j, \Psi_k^{(q)}) dx_j}{\gamma_{ijk}^{(q)}}; \tag{20}$$

$$\xi_{ijk}^{(q)} = \mathbb{E}_{\Psi^{(q)}}[x_{ij}^2 | \tilde{\mathbf{x}}_i, Z_{ik} = 1] = \frac{\displaystyle\int x_j^2 \mu_{\tilde{x}_{ij}}(x_j) g_{jk}(x_j, \Psi_k^{(q)}) dx_j}{\gamma_{ijk}^{(q)}}. \tag{21}$$

With these notations, using Bayes' theorem, we have:

$$t_{ik}^{(q)} = \mathbb{E}_{\Psi^{(q)}}[Z_{ik}|\tilde{\mathbf{x}}_i] = \frac{\mathbb{P}_{\Psi^{(q)}}(\tilde{\mathbf{x}}_i | Z_{ik=1}) \mathbb{P}_{\Psi^{(q)}}(Z_{ik} = 1)}{\mathbb{P}_{\Psi^{(q)}}(\tilde{\mathbf{x}}_i)} = \frac{\gamma_{ik}^{(q)} \pi_k^{(q)}}{p_i^{(q)}}. \tag{22}$$

Now, using assumptions (4) and (15), we get:

$$\begin{aligned} \mathbb{E}_{\Psi^{(q)}} \left[ \sum_{j=1}^{p} \frac{Z_{ik}}{\sigma_{jk}^2} (x_{ij} - m_{jk})^2 |\tilde{\mathbf{x}}_i \right] = & \sum_{j=1}^{p} \frac{1}{\sigma_{jk}^{(q)2}} \Big( \mathbb{E}_{\Psi^{(q)}} \left[ Z_{ik} x_{ij}^2 | \tilde{\mathbf{x}}_i \right] \\ & - 2\, m_{jk}^{(q)}\, \mathbb{E}_{\Psi^{(q)}} \left[ Z_{ik} x_{ij} | \tilde{\mathbf{x}}_i \right] + m_{jk}^{(q)2}\, \mathbb{E}_{\Psi^{(q)}} \left[ Z_{ik} | \tilde{\mathbf{x}}_i \right] \Big). \end{aligned} \tag{23}$$

Furthermore,

$$\mathbb{E}_{\Psi^{(q)}}\left[Z_{ik}x_{ij}^2|\tilde{\mathbf{x}}_i\right] = \mathbb{E}_{\Psi^{(q)}}\left[x_{ij}^2|\tilde{x}_{ij}, Z_{ik}=1\right]\mathbb{P}\left(Z_{ik}=1|\tilde{\mathbf{x}}_i\right) = \xi_{ijk}^{(q)}t_{ik}^{(q)}, \quad (24)$$

$$\mathbb{E}_{\Psi^{(q)}}\left[Z_{ik}x_{ij}|\tilde{\mathbf{x}}_i\right] = \mathbb{E}_{\Psi^{(q)}}\left[x_{ij}|\tilde{x}_{ij}, Z_{ik}=1\right]\mathbb{P}\left(Z_{ik}=1|\tilde{\mathbf{x}}_i\right) = \eta_{ijk}^{(q)}t_{ik}^{(q)}. \quad (25)$$

Hence, finally, Equation (16) becomes:

$$\begin{aligned} Q(\Psi,\Psi^{(q)}) = {} & \sum_{k=1}^{g}\log\pi_k\sum_{i=1}^{n}t_{ik}^{(q)} - \frac{np}{2}\log(2\,\pi) - \sum_{k=1}^{g}\sum_{j=1}^{p}\log\sigma_{jk}\sum_{i=1}^{n}t_{ik}^{(q)} \\ & - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{g}t_{ik}^{(q)}\left(\sum_{j=1}^{p}\frac{1}{\sigma_{jk}^2}\,\xi_{ijk}^{(q)} - 2\sum_{j=1}^{p}\frac{m_{jk}}{\sigma_{jk}^2}\,\eta_{ijk}^{(q)} + \sum_{j=1}^{p}\frac{m_{jk}^2}{\sigma_{jk}^2}\right). \quad (26) \end{aligned}$$

**M-step of the FEM algorithm** In order to maximize $Q(\Psi,\Psi^{(q)})$ defined by Equation (26), its partial derivatives with respect to the various parameters have to be set to zero. The partial derivatives with respect to the proportions $\pi_k$ are:

$$\frac{\partial Q(\Psi,\Psi^{(q)})}{\partial\pi_k} = \frac{1}{\pi_k}\sum_{i=1}^{n}t_{ik}^{(q)};$$

equating these derivatives to zero, under the constraint $\sum_{k=1}^{g}\pi_k = 1$, give similar results to the EM algorithm for crisp data:

$$\pi_k^{(q+1)} = \frac{1}{n}\sum_{i=1}^{n}t_{ik}^{(q)}. \quad (27)$$

Computing derivatives with respect to each element $m_{jk}$ of the means gives:

$$\frac{\partial Q(\Psi,\Psi^{(q)})}{\partial m_{jk}} = \frac{1}{\sigma_{jk}^2}\sum_{i=1}^{n}t_{ik}^{(q)}\eta_{ijk}^{(q)} - m_{jk}\sum_{i=1}^{n}t_{ik}^{(q)}; \quad (28)$$

setting this partial derivative to zero, we get the following update equations:

$$m_{jk}^{(q+1)} = \frac{\sum_{i=1}^{n}t_{ik}^{(q)}\eta_{ijk}^{(q)}}{\sum_{i=1}^{n}t_{ik}^{(q)}}. \quad (29)$$

Eventually, the first-order derivative of $Q(\Psi,\Psi^{(q)})$ with respect to $\sigma_{jk}$ is:

$$\frac{\partial Q(\Psi,\Psi^{(q)})}{\partial\sigma_{jk}} = -\frac{1}{\sigma_{jk}}\sum_{i=1}^{n}t_{ik}^{(q)} + \frac{1}{\sigma_{jk}^3}\sum_{i=1}^{n}t_{ik}^{(q)}\left(\xi_{ijk}^{(q)} - 2\,m_{jk}\eta_{ijk}^{(q)} + m_{jk}^2\right).$$

Setting this partial derivative to zero gives:

$$\sigma_{jk}^{(q+1)} = \sqrt{\frac{\sum_{i=1}^{n}t_{ik}^{(q)}\left(\xi_{ijk}^{(q)} - 2\,m_{jk}^{(q+1)}\eta_{ijk}^{(q)} + m_{jk}^{(q+1)\,2}\right)}{\sum_{i=1}^{n}t_{ik}^{(q)}}}. \quad (30)$$

*Remark 2 (Spherical model).* Assume, as in Remark 1, that the classes are spherical: $\Sigma_k = \sigma_k^2 \mathrm{Id_p}$. Then, the update equations for the proportions and the means are unchanged, and the update equations for the standard deviations become:

$$\sigma_k^{(q+1)} = \sqrt{\frac{\sum_{i=1}^n t_{ik}^{(q)} \sum_{j=1}^p \left( \xi_{ijk}^{(q)} - 2 m_{jk}^{(q+1)} \eta_{ijk}^{(q)} + m_{jk}^{(q+1)2} \right)}{p \sum_{i=1}^n t_{ik}^{(q)}}}. \tag{31}$$

*Remark 3 (Relationship between the update equations for crisp and fuzzy data).* We may notice the similarity with the update equations obtained for crisp data. The difference is that the crisp quantities $x_{ij}$ and $x_{ij}^2$ are replaced with the conditional expectations $\eta_{ijk}$ and $\xi_{ijk}$ of the fuzzy variables $\tilde{x_{ij}}$ and $\tilde{x_{ij}}^2$, respectively.

*Remark 4 (Prior on the covariance matrices).* Suppose that a prior is set on each covariance matrix $\Sigma_k$ using the inverse-Wishart distribution with parameters $m_0$ and $\Lambda_{k\,0} = \mathrm{diag}\,(\lambda_{k1}, \ldots, \lambda_{kp})$. Then, the estimates for the standard deviations for the conditional independence case becomes:

$$\sigma_{jk}^{(q+1)} = \sqrt{\frac{\sum_{i=1}^n t_{ik}^{(q)} \left( \xi_{ijk}^{(q)} - 2\, m_{jk}^{(q+1)} \eta_{ijk}^{(q)} + m_{jk}^{(q+1)2} \right) + \lambda_{jk}}{\sum_{i=1}^n t_{ik}^{(q)} + (m_0 + p + 1)}}. \tag{32}$$

In the spherical case, using $\Lambda_{k\,0} = \lambda_k \mathrm{Id_p}$, we get:

$$\sigma_k^{(q+1)} = \sqrt{\frac{\sum_{i=1}^n t_{ik}^{(q)} \sum_{j=1}^p \left( \xi_{ijk}^{(q)} - 2 m_{jk}^{(q+1)} \eta_{ijk}^{(q)} + m_{jk}^{(q+1)2} \right) + \lambda_k}{p \sum_{i=1}^n t_{ik}^{(q)} + p\,(m_0 + p + 1)}}. \tag{33}$$

# 4 Experiments

## 4.1 Synthetic data

First, we ran experiments over synthetic two-dimensional data. We placed ourselves in the experimental setting considered in [8]: here, a fuzzy datum represents the imprecise knowledge of the actual (precise) value of a variable. We generated data as follows. First, we drew a sample of $n = 300$ realizations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of a Gaussian mixture of $g = 3$ components with the parameters given in Table 1. Level curves of the corresponding density are represented in Figure 1. The curves correspond to levels 0.01, 0.02, 0.03, 0.04 and 0.05 of the density.

Each data point was classified according to the Bayes' rule. Then, each $\mathbf{x}_i$ was transformed into a fuzzy number. Let a (monovariate) trapezoidal fuzzy number $\tilde{w}$ be defined by four scalars $a$, $b$, $c$ and $d$, such that:

$$\mu_{\tilde{w}}(w) = \begin{cases} (w-b)/(b-a), & \text{if } a \leq w \leq b, \\ 1, & \text{if } b \leq w \leq c, \\ (d-w)/(d-c), & \text{if } c \leq w \leq d, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

**Table 1.** Parameters of the components of the Gaussian mixture.

| | comp. 1 | comp. 2 | comp. 3 |
|---|---|---|---|
| $\pi_k$ | 0.3 | 0.4 | 0.3 |
| $m_k$ | $(-2,-2)^\top$ | $(-1,+1)^\top$ | $(+2,-2)^\top$ |
| $\Sigma_k$ | $\begin{pmatrix} 2.5 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ | $\begin{pmatrix} 2 & 0 \\ 0 & 1.5 \end{pmatrix}$ | $\begin{pmatrix} 1.25 & -0.25 \\ -0.25 & 1 \end{pmatrix}$ |

Here, each coordinate $x_{ij}$ of a crisp data point $\mathbf{x}_i$ was transformed into a trapezoidal fuzzy number $\tilde{x_{ij}}$ as follows. Four iid realizations $u_1$, $u_2$, $u_3$, and $u_4$ were drawn from an uniform distribution $\mathcal{U}_{[0;1]}$; then, $r$ and $s$ being user-defined:

$$\begin{aligned} b_{ij} = u_1(x_{ij} - r),\ \ a_{ij} = u_2(b_{ij} - s); \\ c_{ij} = u_3(x_{ij} + r),\ \ d_{ij} = u_4(x_{ij} + s). \end{aligned} \qquad (35)$$

Figure 1 displays the fuzzy numbers thus obtained using $r = 0.5$ and $s = 2$: each rectangle corresponds to the alpha-cut of a membership function, with $\alpha = 0.75$. The line style of each rectangle (plain, dashed, or dotted) represents the class with highest probability, determined using the true distribution of the data.

**Table 2.** Parameter estimates, number $q$ of iterations, log-likelihood, and PRI obtained with $r = 0.5$ and $s = 2$.

| | | $\pi_k$ | $m_k$ | $\Sigma_k$ | $q$ | $\log L$ | PRI |
|---|---|---|---|---|---|---|---|
| | comp. 1 | 0.41 | $(-2,-1.58)^\top$ | diag (1.37, 1.37) | | | |
| spherical | comp. 2 | 0.28 | $(-0.75, 1.49)^\top$ | diag (1.06, 1.06) | 18 | -969.1 | 0.7984 |
| | comp. 3 | 0.31 | $(1.74,-2.04)^\top$ | diag (0.75, 0.75) | | | |
| | comp. 1 | 0.31 | $(-1.94,-2.15)^\top$ | diag (1.78, 0.39) | | | |
| diagonal | comp. 2 | 0.41 | $(-1, 0.94)^\top$ | diag (1.59, 1.53) | 68 | -963.8 | 0.9136 |
| | comp. 3 | 0.28 | $(1.82,-2.11)^\top$ | diag (0.79, 0.60) | | | |

Parameters were estimated using the FEM algorithm, for the two models studied in Section 3.2. From now on, the model with diagonal covariance matrices will be referred to as diagonal model. The means were initialized at random, according to a centered and scaled normal distribution; the initial covariance matrices were set to the identity. The quality of each partition was evaluated using the pairwise Rand index (PRI). For each model, we performed $N = 20$ runs of the algorithm; we retained the best result (for which the log-likelihood was maximal). Parameter estimates are given in Table 2, as well as the corresponding value of the fuzzy log-likelihood, the number $q$ of iterations, and the PRI. Figure 2 displays the densities estimated by both models, at the same levels as previously. The fuzzy numbers are also represented; now, the line style of each rectangle indicates the class into which the corresponding data point was classified.

Then, we modified the synthetic dataset as follows. We fuzzified the same crisp data using values $r = 2$ and $s = 2$. Figure 3 represents the alpha-cuts (with $\alpha = 0.75$) of the fuzzy numbers thus obtained. As previously, the algorithms were run $N = 20$ times using the same initialization procedure, and the best result was retained for each model. The results are given in Table 3. Figure 3 also displays the density estimated using the diagonal model. The estimates of the means are very similar to the previous; those of the variances, however, are much lower. Indeed, the imprecision on the actual realizations of the random variables is higher. In other terms, the intervals in which these realizations may fall with the same degree of possibility as previously are larger. Then, the algorithm obviously favours a solution with as small variances as possible, as it maximizes the likelihood. Finally, we set an inverse-Wishart prior on the covariance matrices. Table 4 presents the results obtained with $m_0 = 2$, and $\Lambda_0 = \mathrm{diag}\,(2\ 2)$ for both models.

**Table 3.** Parameter estimates, number $q$ of iterations, log-likelihood, and PRI obtained with $r = 2$ and $s = 2$.

| | $\pi_k$ | $m_k$ | $\Sigma_k$ | $q$ | $\log L$ | PRI |
|---|---|---|---|---|---|---|
| comp. 1 | 0.49 | $(-1.71, -1.26)^\top$ | $\mathrm{diag}\,(1.29, 1.29)$ | | | |
| spherical comp. 2 | 0.21 | $(-0.63, 1.73)^\top$ | $\mathrm{diag}\,(0.59, 0.59)$ | 165 | -587.7 | 0.6937 |
| comp. 3 | 0.30 | $(1.70, -1.90)^\top$ | $\mathrm{diag}\,(0.63, 0.63)$ | | | |
| comp. 1 | 0.53 | $(-1.27, -1.50)^\top$ | $\mathrm{diag}\,(1.99, 0.73)$ | | | |
| diagonal comp. 2 | 0.25 | $(-0.86, 1.63)^\top$ | $\mathrm{diag}\,(1.05, 0.48)$ | 209 | -585 | 0.6536 |
| comp. 3 | 0.22 | $(2.02, -2.06)^\top$ | $\mathrm{diag}\,(0.27, 0.73)$ | | | |

**Table 4.** Parameter estimates, number $q$ of iterations, log-likelihood, and PRI obtained with $r = 2$ and $s = 2$, with an inverse-Wishart prior on the covariance matrices.

| | $\pi_k$ | $m_k$ | $\Sigma_k$ | $q$ | $\log L$ | PRI |
|---|---|---|---|---|---|---|
| comp. 1 | 0.37 | $(-2.16, -1.54)^\top$ | $\mathrm{diag}\,(1.14, 1.14)$ | | | |
| spherical comp. 2 | 0.31 | $(0.78, 1.16)^\top$ | $\mathrm{diag}\,(1.18, 1.18)$ | 95 | -622.1 | 0.7913 |
| comp. 3 | 0.32 | $(1.47, -1.97)^\top$ | $\mathrm{diag}\,(0.83, 0.83)$ | | | |
| comp. 1 | 0.34 | $(-2.19, -1.62)^\top$ | $\mathrm{diag}\,(1.27, 0.95)$ | | | |
| diagonal comp. 2 | 0.33 | $(-0.82, 1.09)^\top$ | $\mathrm{diag}\,(1.23, 1.19)$ | 331 | -620.9 | 0.8427 |
| comp. 3 | 0.33 | $(1.38, -2)^\top$ | $\mathrm{diag}\,(1.11, 0.64)$ | | | |

### 4.2 Real data

We ran the FEM algorithm on the blood data used in [6]. This dataset presents statistics on daily measurements of systolic and diastolic pressures on $n = 108$

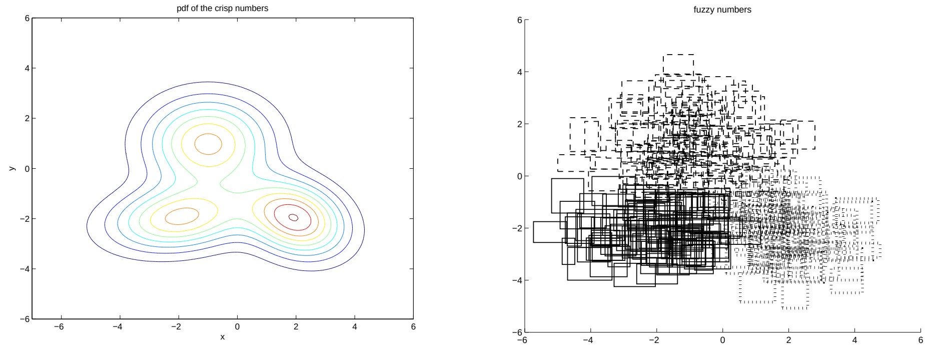


**Fig. 1.** Pdf of the Gaussian mixture (left); fuzzy numbers obtained with $r = 0.5$ and $s = 2$ (right).

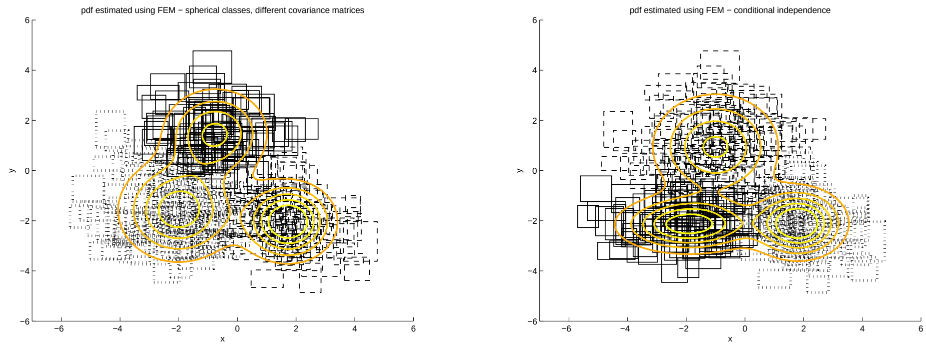


**Fig. 2.** Pdf estimated using the spherical model (left) and the diagonal model (right).

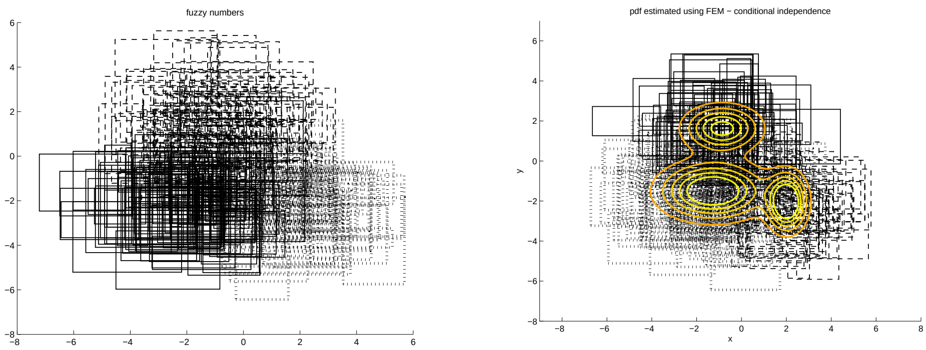


**Fig. 3.** Fuzzy numbers obtained with $r = 2$ and $s = 2$ (left); density estimated using the diagonal model (right).

patients. Remark that here, each measurement is precise; however, for each patient, center and spread values only were stored. Thus, the fuzziness of a datum stems from the variability of the measurements performed on each patient and the choice to summarize these measurements using their center and their spread only. Although this interpretation differs from the point of view adopted in this paper, we used these data to compare our results with those presented in [6].

We interpreted these data as triangular fuzzy numbers, which are a special case for trapezoidal fuzzy numbers. In addition, we assumed that each center was equidistant to the minimal and maximal values. First, the data were centered and scaled with respect to the mean and standard deviation of the center values. Then, the density of a two-component mixture of Gaussians was estimated using the FEM algorithm. The parameters were initialized as previously.

The alpha-cuts of the fuzzy numbers are represented in Figure 4 (again, with $\alpha = 0.75$). Table 5 presents the results obtained with both models using an inverse-Wishart prior on $\Sigma_1$ and $\Sigma_2$, with $m_0 = 2$ and $\Lambda_{01} = \Lambda_{02} = \mathrm{diag}\,(2, 2)$. Figure 4 displays the density estimated using the diagonal model, at levels 0.02, 0.04, 0.06, 0.08 and 0.1. The results are quite similar to those obtained in [6]. Here, 55 patients were assigned to the first class, and 53 to the second one.

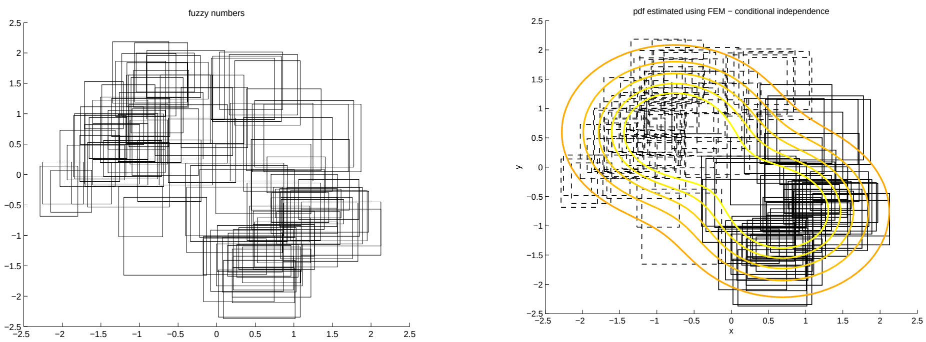


**Fig. 4.** Fuzzy numbers obtained from the blood data (left); density estimated using the diagonal model with an inverse-Wishart prior on the covariance matrices (right).

## 5 Conclusion

In this paper, we addressed the problem of clustering fuzzy data using mixture models. Our approach is based on an extention of the EM algorithm for fuzzy data, proposed by Denœux [4, 5]. The likelihood of a mixture of Gaussians may be computed, given a sample of fuzzy numbers, using Zadeh's extention principle. Then, the estimates maximizing this likelihood may be estimated using an iterative procedure. At each iteration, the expectation $Q(\Psi, \Psi^{(q)}) = \mathbb{E}[\log L_c(\Psi)|\tilde{\mathbf{y}}]$

**Table 5.** Parameter estimates, number $q$ of iterations, and log-likelihood obtained for the blood data, with an inverse-Wishart prior on the covariance matrices.

| | $\pi_k$ | $m_k$ | $\Sigma_k$ | $q$ | $\log L$ |
|---|---|---|---|---|---|
| spherical comp. 1 | 0.523 | $(-0.768, 0.599)^\top$ | diag $(0.558, 0.558)$ | 66 | -220.4 |
| comp. 2 | 0.477 | $(0.690, -0.757)^\top$ | diag $(0.532, 0.532)$ | | |
| diagonal comp. 1 | 0.523 | $(-0.767, 0.598)^\top$ | diag $(0.567, 0.549)$ | 32 | -220.4 |
| comp. 2 | 0.477 | $(0.691, -0.757)^\top$ | diag $(0.520, 0.543)$ | | |

of the log-likelihood is first computed. Then, the parameters of the model may be updated so as to maximize this expectation. In this paper, we detailed the computation of the update equations under the assumption that the covariance matrices considered are diagonal, in the case of a finite mixture of Gaussians.

We conducted experiments on synthetic and real data. Experiments show that our algorithm estimates accurately the distribution of imprecisely known data. Our approach may be sensitive to the amount of imprecision in the available information. In particular, the covariance matrices may be under-estimated if the degree of fuzziness is high. However, the algorithm may be guided towards a desired solution by setting a prior distribution on these parameters. Thus, our algorithm constitutes a generic approach for clustering imprecise data.

The extension of the fuzzy EM algorithm to finite mixture of Gaussians with full covariance matrices is straightforward. However, in such cases, it may be necessary to rely on Monte Carlo processes in order to perform the E-step of the EM algorithm. Therefore, this work is left for further research.

## References


1. Coppi, R., D'Urso, P.: Fuzzy K-means clustering models for triangular fuzzy time trajectories. Statistical Methods and Applications 11 (1), pages 21-40 (2002).
2. Coppi, R., D'Urso, P.: Three-way fuzzy clustering models for LR fuzzy time trajectories. Computational Statistics and Data Analysis 43, pages 149-177 (2003).
3. Dempster, A. P., Laird, N. M. , Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, pages 1-38 (1977).
4. Denœux, T.: Maximum likelihood estimation from evidential data. In Proceedings of the first workshop on the theory of belief functions and their applications, Brest, France (2010). Manuscript available online at http://www.ensieta.fr/belief2010/.
5. Denœux, T.: Maximum likelihood estimation from fuzzy data using the Fuzzy EM algorithm (working paper).
6. D'Urso, P., Giordani, P.: A weighted fuzzy $c$-means clustering model for fuzzy data. Computational Statistics and Data Analysis 50, pages 1496-1523 (2006).
7. Auephanwiriyakul, S., Keller, J.M.: Analysis and efficient implementation of a linguistic fuzzy c-means. IEEE Trans. on Fuzzy Systems 10 (5), pages 563-582 (2002).
8. Hamdan, H., Govaert, G.: CEM algorithm for imprecise data. Application to flaw diagnosis using acoustic emission. In Proc. of the IEEE International conference on Systems, Man and Cybernetics (5), pages 4774-4779 (The Hague, Netherlands, 2004).

9. Hamdan, H., Govaert, G.: Mixture model clustering of uncertain data. In Proc. of the IEEE International Conference on Fuzzy Systems, pages 879-884 (Reno, Nevada, USA, 2005).
10. Hathaway, R.J., Bezdek, J.C., Pedrycz,W.: A parametric model for fusing heterogeneous fuzzy data. IEEE Trans. on Fuzzy Systems 4 (3), pages 1277-1282 (1996).
11. Hung, W.-L., Yang, M.-S.: Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. Fuzzy Sets and Systems 150 (3), pages 561-577 (2005).
12. Gebhardt, J., Gil, M. A., Kruse, R.: Fuzzy set-theoretic methods in statistics. In R. Slowinski, editor, Fuzzy sets in decision analysis, operations research and statistics, pages 311-347. Kluwer Academic Publishers, Boston (1998).
13. Gil, M. A., López-Díaz, M., Ralescu, D. A.: Overview on the development of fuzzy random variables. Fuzzy Sets and Systems, 157(19):2546–2557 (2006).
14. Jordan, M., Jacobs, R.: Hierarchical mixtures of experts and the EM algorithm. Neural Computation 6, pages 181-214 (1994).
15. Kruse, R., Meyer, K. D.: Statistics with vague data. Kluwer, Dordrecht (1987).
16. Mauris, G.: Expression of measurement uncertainty in a very limited knowledge context: a possibility theory-based approach. IEEE Trans. on Instrumentation and Measurement 56 (3), pp. 731-735 (2007).
17. McLachlan, G. J. , Krishnan, T.: The EM Algorithm and Extensions. Wiley, New York (1997).
18. Pelekis, N., Iakovidis, D., Kotsifakos, E., Kopanakis, I.: Fuzzy Clustering of Intuitionistic Fuzzy Data. International Journal of Business Intelligence and Data Mining 3 (1), pages 45-65 (2008).
19. Redner, R., Walker, H.: Mixture densities, maximum likelihood and the EM algorithm. SIAM review 26 (2), pages 195-239 (1984).
20. Sato, M., Sato,Y.: Fuzzy clustering model for fuzzy data. In Proc. of the 4th IEEE Conf. on Fuzzy Systems, pages 2123-2128, Yokohama, Japan (1995).
21. Takata, O., Miyamoto, S., Umayahara, K.: Clustering of data with uncertainties using Hausdorff distance. In Proc. of the 2nd IEEE International Conference on Intelligence Processing Systems, pages 67-71 (Gold Coast, Australia, 1998).
22. Takata, O., Miyamoto, S., Umayahara, K.: Fuzzy clustering of data with uncertainties using minimum and maximum distances based on L1 metric. In Proc of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference, pages 2511-2516 (Vancouver, British Columbia, Canada, 2001).
23. Viertl, R.: Univariate statistical analysis with fuzzy data. Computational Statistics & Data Analysis, 51(1):133–147 (2006).
24. Wu, C.F.J.: On the convergence properties of the EM algorithm. Annals of Statistics 11, pages 95-103 (1983).
25. Xu, Z., Chen, J., WU, J.: Clustering algorithm for intuitionistic fuzzy sets. Information Sciences 178, pages 3775-3790 (2008).
26. Yang, M.S., Ko, C.H.: On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets and Systems 84, pages 49-60 (1996).
27. Yang, M.S., Liu, H.H.: Fuzzy clustering procedures for conical fuzzy vector data. Fuzzy Sets and Systems 106, pages 189-200 (1999).
28. Yang, M.S., Hwang, P.Y., Chen, D.H.: Fuzzy clustering algorithms for mixed feature variables. Fuzzy Sets and Systems 141, pages 301-317 (2004).
29. Zadeh, L. A.: Fuzzy sets. Information and Control 8, pages 338-353 (1965).
30. Zadeh, L. A.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications 10, pages 421-427 (1968).