Mixture model of Gaussian copulas to cluster mixed-type data

Matthieu Marbac, *DGA & Inria Lille & University Lille 1*,
matthieu.marbac-lourdelle@inria.fr
Christophe Biernacki, *University Lille 1 & CNRS & Inria Lille*,
christophe.biernacki@math.univ-lille1.fr
Vincent Vandewalle, *EA 2694 University Lille 2 & Inria Lille*,
vincent.vandewalle@univ-lille2.fr

Abstract. A mixture model of Gaussian copulas is proposed to cluster mixed data. This approach allows to straightforwardly define simple multivariate intra-class dependency models while preserving classical distributions for the one-dimensional margins of each component in order to facilitate the model interpretation. Moreover, the intra-class dependencies are taken into account by the Gaussian copulas which provide one robust correlation coefficient per couple of variables and per class. This model generalizes different existing models defined for homogeneous or mixed variables. The Bayesian inference is performed via a Metropolis-within-Gibbs sampler. The model is illustrated by a real data set clustering.

Keywords. Clustering, Gaussian copula, Gibbs sampler, Mixed data, Mixture models.

1 Introduction

With the informatics advent, multivariate data sets become more complex. Particularly, they often contain mixed data (variables of different kinds). *Clustering* provides an efficient solution to extract the main information from the data by grouping the individuals into few characteristic classes. It can be performed by probabilistic methods modelling the data generation whose the most popular one uses finite mixture models of parametric components [12]. In such a case, a class gathers together the individuals drawn by the same distribution. Obviously, the choice of the component distributions depends on the kind of the variables at hand. However, few distributions exist to model mixed data and their margin distributions are often complex [8].

The simplest way to cluster mixed variables consists in approaching the data distribution with a finite mixture model assuming independence conditionally on the class membership of each individual. This model, called *locally independent model*, obtains good results in many real clustering problems [11, 6], especially when few individuals are described by several variables.

Indeed, when its one-dimensional margins of each component follow classical distributions, this model provides a meaningful summary of the data by its margin parameters. However, this model leads to biases when its assumption of conditional independence is violated.

The aim of this paper is to present a model-based clustering for mixed data of any kinds of variables admitting a cumulative distribution function. This model has a double objective: to preserve *classical distributions* for *all* its margin distributions of each component and to *model the intra-class dependencies*. This objective can naturally be achieved by the use of copulas [9] since these objects allow to build a multivariate model by setting, on the one hand, the one-dimensional *margins*, and, on the other hand, the *dependency model* between variables. More precisely, the data distribution is approached by a full parametric *mixture model of Gaussian copulas* whose the margin distributions of each component are classical and whose the Gaussian copulas [7] model the intra-class dependencies. The new mixture model is meaningful since each class is summarized by its proportion, by the parameters of each marginal distributions and by the correlation matrix of the Gaussian copula providing one coefficient per couple of variables measuring the intra-class dependency. In addition, a principal component analysis (PCA) computed per class is a straightforward by-product of the model. Indeed, it is computed on the correlation matrix of the class and it can be used to summarize the main intra-class dependencies and to provide a scatter-plot of the individuals according to the class parameters.

This paper is organized as follows. Section 2 presents the mixture model of Gaussian copulas for clustering, its links with the existing models and its contribution to the visualization of mixed variables. Section 3 is devoted to the parameter estimation in a Bayesian framework. Section 4 illustrates the model by a real data set clustering. Section 5 concludes this work.

2 Mixture model of Gaussian copulas

Finite mixture model

Let the vector of e mixed variables $\boldsymbol{x} = (x^1, \ldots, x^e) \in \mathbb{R}^c \times \mathcal{X}$, whose the first c elements are the set of the continuous variables further denoted by $\boldsymbol{x}^{\mathrm{C}}$, and whose the last d elements are the set of the discrete variables (integer, ordinal or binary) further denoted by $\boldsymbol{x}^{\mathrm{D}}$, with $e = c + d$. Note that if x^j is an ordinal variable with m_j modalities, then it uses a numeric coding $\{1, \ldots, m_j\}$. Data $\boldsymbol{x}$ are supposed to be drawn by the mixture model of g parametric distributions whose the probability distribution function (pdf) is written as

$$p(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{k=1}^{g} \pi_k p(\boldsymbol{x}; \boldsymbol{\alpha}_k), \tag{1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\alpha})$ and where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_g)$ groups the proportions of each class k denoted by π_k, and respects the following constraints $0 < \pi_k \leq 1$ and $\sum_{k=1}^{g} \pi_k = 1$, while $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_g)$ groups the parameters of each class k denoted by $\boldsymbol{\alpha}_k$.

One-dimensional margins of the components

The margin distribution of x^j, for the component k, belongs to the exponential family and has $p(x^j; \boldsymbol{\beta}_{kj})$ for pdf and $P(x^j; \boldsymbol{\beta}_{kj})$ for cumulative distribution function (cdf). More precisely, the margin distribution of each component is a *Gaussian* (if x^j is *continuous*), *Poisson* (if x^j is *integer*) or *multinomial* (if x^j is *ordinal*) distribution where $\boldsymbol{\beta}_{kj}$ denotes the usual parameters.

Dependency model of the components

The model assumes that each component k follows a Gaussian copula whose the correlation matrix is $\mathbf{\Gamma}_k$. We note $\Phi_e(.;\mathbf{\Gamma}_k)$ the cdf of the e-variate centred Gaussian distribution with correlation matrix $\mathbf{\Gamma}_k$, and $\Phi_1^{-1}(.)$ the inverse cumulative distribution function of univariate Gaussian variable $\mathcal{N}_1(0,1)$. Thus, the cdf of the component k is written as

$$P(\boldsymbol{x};\boldsymbol{\alpha}_k) = \Phi_e(\Phi_1^{-1}(u_k^1),\ldots,\Phi_1^{-1}(u_k^e);\mathbf{0},\mathbf{\Gamma}_k), \tag{2}$$

where $u_k^j = P(x^j;\boldsymbol{\beta}_{kj})$, $\boldsymbol{\alpha}_k = (\boldsymbol{\beta}_k,\mathbf{\Gamma}_k)$ and $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_{k1},\ldots,\boldsymbol{\beta}_{ke})$.

Remark 2.1 (Standardized coefficient of correlation per class)**.**
The Gaussian copula provides a robust coefficient of correlation per couple of variables. Indeed, when both variables are continuous, it is equal to the upper bound of the coefficient of correlation obtained by all the monotonic transformations of the variables [10]. Furthermore, when both variables are discrete, it is equal to the polychoric coefficient of correlation [13].

Remark 2.2 (Two latent variables)**.**
The mixture model of Gaussian copulas involves two latent variables: a categorical one using a condense coding $z \in \{1,\ldots,g\}$ denoting the class membership and an e-variate Gaussian one $\boldsymbol{y} = (y^1,\ldots,y^e) \in \mathbb{R}^e$. Indeed, if $\boldsymbol{y}|z = k \sim \mathcal{N}_e(\mathbf{0},\mathbf{\Gamma}_k)$ and if $x^j = P^{-1}(\Phi_1(y^j);\boldsymbol{\beta}_{kj})$, $\forall j = 1\ldots,e$, then the component k is a Gaussian copula whose the cdf is defined in (2)*. Thus, we deduce the following generative model*

- *Class membership* sampling*: $z \sim \mathcal{M}_g(\pi_1,\ldots,\pi_g)$*
- *Gaussian copula* sampling*: $\boldsymbol{y}|z = k \sim \mathcal{N}_e(\mathbf{0},\mathbf{\Gamma}_k)$*
- *Observed data* deterministic computation *of $\boldsymbol{x}$ as such $x^j = P^{-1}(\Phi_1(y^j);\boldsymbol{\beta}_{kj})$.*

Probability distribution function of the components

We introduce the function $\Psi(\boldsymbol{x}^{\mathrm{C}};\boldsymbol{\alpha}_k) = \left(\frac{x^j-\mu_{kj}}{\sigma_{kj}}; j = 1,\ldots,c\right)$ and the space of the antecedents of $\boldsymbol{x}^{\mathrm{D}}$ in the class k, by $\mathcal{S}_k = \mathcal{S}_k^{c+1} \times \ldots \times \mathcal{S}_k^e$, where $\mathcal{S}_k^j$ is the interval defined by $\mathcal{S}_k^j =]b_k^{\ominus}(x^j), b_k^{\oplus}(x^j)]$, for $j = c+1,\ldots,e$, whose the bounds are $b_k^{\ominus}(x^j) = \Phi_1^{-1}(P(x^j-1;\boldsymbol{\beta}_{kj}))$ and $b_k^{\oplus}(x^j) = \Phi_1^{-1}(P(x^j;\boldsymbol{\beta}_{kj}))$. The pdf of the component k is written as

$$p(\boldsymbol{x};\boldsymbol{\alpha}_k) = p(\boldsymbol{x}^{\mathrm{C}};\boldsymbol{\alpha}_k)p(\boldsymbol{x}^{\mathrm{D}}|\boldsymbol{x}^{\mathrm{C}};\boldsymbol{\alpha}_k) \tag{3}$$

$$= \frac{\phi_c(\Psi(\boldsymbol{x}^{\mathrm{C}};\boldsymbol{\alpha}_k);\mathbf{0},\mathbf{\Gamma}_{k\mathrm{CC}})}{\prod_{j=1}^c \sigma_{kj}} \int_{\mathcal{S}_k} \phi_d(\boldsymbol{u};\boldsymbol{\mu}_k^{\mathrm{D}},\mathbf{\Sigma}_k^{\mathrm{D}})d\boldsymbol{u}, \tag{4}$$

where $\mathbf{\Gamma}_k = \begin{bmatrix} \mathbf{\Gamma}_{k\mathrm{CC}} & \mathbf{\Gamma}_{k\mathrm{CD}} \\ \mathbf{\Gamma}_{k\mathrm{DC}} & \mathbf{\Gamma}_{k\mathrm{DD}} \end{bmatrix}$ is decomposed into sub-matrices, for instance $\mathbf{\Gamma}_{k\mathrm{CC}}$ is the sub-matrix of the first c rows and columns of $\mathbf{\Gamma}_k$, where $\boldsymbol{\mu}_k^{\mathrm{D}} = \mathbf{\Gamma}_{k\mathrm{DC}}\mathbf{\Gamma}_{k\mathrm{CC}}^{-1}\Psi(\boldsymbol{x}^{\mathrm{C}};\boldsymbol{\alpha}_k)$ is the conditional mean of $\boldsymbol{y}^{\mathrm{D}}$ and where $\mathbf{\Sigma}_k^{\mathrm{D}} = \mathbf{\Gamma}_{k\mathrm{DD}} - \mathbf{\Gamma}_{k\mathrm{DC}}\mathbf{\Gamma}_{k\mathrm{CC}}^{-1}\mathbf{\Gamma}_{k\mathrm{CD}}$ is its conditional covariance matrix.

Heteroscedastic and homoscedastic versions of the model

The trade off between the bias and the variance of the model may be improved by adding some constraints on the parameter space. Thus, we propose an homoscedastic version of the mixture model of Gaussian copulas by assuming the equality between the correlation matrices, so

$$\mathbf{\Gamma}_1 = \ldots = \mathbf{\Gamma}_g. \tag{5}$$

The heteroscedastic (resp. homoscedastic) mixture model of Gaussian copulas requires ν_{He} (respectively ν_{Ho}) parameters where

$$\nu_{\text{He}} = (g-1) + g\left(\frac{e(e+1)}{2} + d\right) \text{ and } \nu_{\text{Ho}} = (g-1) + \frac{e(e-1)}{2} + g(e+d). \tag{6}$$

Related models

The mixture model of Gaussian copulas allows to generalize many classical mixture models, among them one can cite the four followers.

- Obviously, if the correlation matrices are diagonal (*i.e.* $\mathbf{\Gamma}_k = \boldsymbol{I}$, $\forall k = 1, \ldots, g$), then the mixture model of Gaussian copulas is equivalent to the *locally independent mixture model*.
- If all the variables are continuous (*i.e.* $c = e$ and $d = 0$), then both versions of the heteroscedastic and homoscedastic mixture models of Gaussian copulas are equivalent to the heteroscedastic and homoscedastic *multivariate Gaussian mixture models* [1].
- The mixture model of Gaussian copulas is linked to the *binned Gaussian mixture model*. For instance, it is equivalent, when data are ordinal, to the mixture model of [5]. In such a case and under the true model assumption, this model is stable by fusion of modalities.
- When the variables are continuous and ordinal, the mixture model of Gaussian copulas is a new parametrization of *model proposed by Everitt* [4] which directly estimates the space S_k containing the antecedents of $\boldsymbol{x}^{\text{D}}$ and not the margin parameters. The maximum likelihood inference is performed via a simplex algorithm dramatically limiting the number of ordinal variables. Note that our approach detailed in Section 3 avoids this drawback.

Data visualization per class: a by-product of Gaussian copulas

We can use the model parameters to perform a *visualization* of the individuals *per class* and to bring out the main intra-class dependencies. Thus, for the class k, we firstly compute the coordinates $\mathbb{E}[\boldsymbol{y}|\boldsymbol{x}, z = k; \boldsymbol{\alpha}_k]$ and we secondly project them on the principal component analysis space of the Gaussian copula of the component k, obtained by the spectral decomposition of $\mathbf{\Gamma}_k$. The individuals drawn by the component k follow a centred Gaussian distribution in the factorial map (so they are close to the origin) while the other ones have an expectation different to zero (so they are farther from the origin). Finally, the correlation circle summarizes the intra-class correlations. The application given in Section 4 illustrates this phenomenon.

3 Bayesian inference

We observe a sample $\mathbf{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ composed by n individuals $\boldsymbol{x}_i \in \mathbb{R}^c \times \mathcal{X}$ assumed to be independently drawn by a mixture model of Gaussian copulas. We assume the independence between the prior distributions and we select the classical conjugate prior distributions for each parameters. The following Gibbs sampler allows to perform the inference, in a Bayesian framework, since its stationary distribution is $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$. Thus, it samples a sequel of parameters according to the marginal posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$. This algorithm relies on two instrumental variables: the class membership of the individuals of $\mathbf{x}$ denoted by $\mathbf{z} = (z_1, \ldots, z_n)$ and the Gaussian vector of the individuals denoted by $\mathbf{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

Algorithm 3.1 (The Gibbs sampler)**.**
Starting from an initial value $\boldsymbol{\theta}^{(0)}$*, its iteration* (r) *is written as*

$$\boldsymbol{z}^{(r)}, \boldsymbol{y}^{(r-1/2)} \sim \boldsymbol{z}, \boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}^{(r-1)} \tag{7}$$

$$\boldsymbol{\beta}_{kj}^{(r)}, \boldsymbol{y}_{[rk]}^{j(r)} \sim \boldsymbol{\beta}_{kj}, \boldsymbol{y}_{[rk]}^{j} | \boldsymbol{x}, \boldsymbol{y}_{[rk]}^{\uparrow j(r)}, \boldsymbol{z}^{(r)}, \boldsymbol{\beta}_{k}^{\uparrow j(r)}, \boldsymbol{\Gamma}_{k}^{(r-1)} \tag{8}$$

$$\boldsymbol{\pi}^{(r)} \sim \boldsymbol{\pi} | \boldsymbol{z}^{(r)} \tag{9}$$

$$\boldsymbol{\Gamma}_{k}^{(r)} \sim \boldsymbol{\Gamma}_{k} | \boldsymbol{y}^{(r)}, \boldsymbol{z}^{(r)}, \tag{10}$$

where $\boldsymbol{y}_{[rk]} = \boldsymbol{y}_{\{i:z_i^{(r)}=k\}}$, $\boldsymbol{y}_i^{\uparrow j(r)} = (y_i^{1(r)}, \ldots, y_i^{j-1(r)}, y_i^{j+1(r-1/2)}, \ldots, y_i^{e(r-1/2)})$ and $\boldsymbol{\beta}_k^{\uparrow j(r)} = (\boldsymbol{\beta}_{k1}^{(r)}, \ldots, \boldsymbol{\beta}_{kj-1}^{(r)}, \boldsymbol{\beta}_{kj+1}^{(r-1)}, \ldots, \boldsymbol{\beta}_{ke}^{(r-1)})$.

Remark 3.2 (Twice sampling of the Gaussian variable)**.**
The Gaussian variable $\boldsymbol{y}$ *is twice generated during one iteration of the Gibbs sampler but, obviously, its stationary distribution stays unchanged. This twice sampling is mandatory because of the strong dependency between* $\boldsymbol{y}$ *and* $\boldsymbol{z}$*, and between* $\boldsymbol{y}_{[rk]}^{j}$ *and* $\boldsymbol{\beta}_{kj}$*.*

Remark 3.3 (On the Metropolis-within-Gibbs sampler)**.**
If the samplings from (9) *and* (10) *are classical, the two other ones are more complex. Indeed, the sampling from* (7) *involves to compute the conditional probabilities of the class memberships, so to compute the integral defined in* (4)*. If the number of discrete variables is large, this computation is time consuming. However, the sampling from* (7) *can be efficiently performed by one iteration of a Metropolis-Hastings algorithm having* $p(z_i, \boldsymbol{y}_i | \boldsymbol{x}_i, \boldsymbol{\theta}^{(r-1)})$ *as stationary distribution. Concerning the sampling according to* (8)*, it is performed in two steps. Firstly, the margin parameter is sampled by one iteration of a Metropolis-Hastings algorithm having* $p(\boldsymbol{\beta}_{kj} | \boldsymbol{x}, \boldsymbol{y}_{[rk]}^{\uparrow j(r)}, \boldsymbol{z}^{(r)}, \boldsymbol{\beta}_{k}^{\uparrow j(r)}, \boldsymbol{\Gamma}_{k})$ *as stationary distribution. Secondly, the latent Gaussian vector is sampled from its full conditional distribution.*

Remark 3.4 (Initialization of the algorithm)**.**
The algorithm is initialized on the maximum likelihood estimate of the locally independent model. Thus, it is initialized in a point close to the maximum of the posterior distribution if the variables are not strongly intra-class correlated.

4 Application: clustering of Portuguese wines

The data The data set [3] contains 6497 variants of the Portuguese "Vinho Verde" wine (1599 red wines and 4898 white wines) described by eleven physiochemical continuous variables (fixed acidity, volatile acidity, citric acidity, residual sugar, chlorides, free sulfur dioxide, total density dioxide, density, pH, sulphates, alcohol) and one integer variable (quality of the wine evaluated by experts). The kinds of the wines (red or white) are hidden and we cluster the data set by excluding of the study one white wine (number 4381) since it is an outlier.

Model selection We estimate the three mixture models (locally independent one, the heteroscedastic and homoscedastic versions of the mixture model of Gaussian copulas) for different numbers of classes. The estimate is obtained by taking the mean of the sampled parameters computed after 1000 iterations. The model selection is performed by using two information criteria (BIC criterion [14], ICL criterion [2]) computed on the maximum *a posteriori* estimate.

We present the values of both used information criteria in Table 1 which distinctly select the bi-component heteroscedastic mixture model of Gaussian copulas.

	g	1	2	3	4	5	6
BIC	loc. indpt.	-63516	-61069	-61010	**-55967**	-60250	-57163
	hetero.	-44675	**-34520**	-39724	-44692	-44484	-48349
	homo.	-44675	-39372	**-38289**	-45209	-43217	-42417
ICL	loc. indpt.	-63516	-61229	-61365	**-56310**	-60726	-58138
	hetero.	-44675	**-34688**	-40176	-44933	-44758	-48959
	homo.	-44675	-39607	**-38791**	-45380	-43345	-42667

Table 1: Values of the BIC and ICL criteria for the three mixture models estimated.

Partition comparison Table 2 presents the values of the adjusted Rand index and the confusion matrices in order to compare the relevance of the estimated partitions according to the true one (wine color). These results confirm that the bi-component heteroscedastic Gaussian copula mixture model is the best one among the competing models since its partition is the closest to the true one.

	white	red
class 1	4359	9
class 2	538	1590

(a) Adj. Rand.: 0.68

	white	red
class 1	2441	12
class 2	1911	7
class 3	545	1580

(b) Adj. Rand.: 0.30

	white	red
class 1	2547	1561
class 2	2007	35
class 3	275	3
class 4	68	0

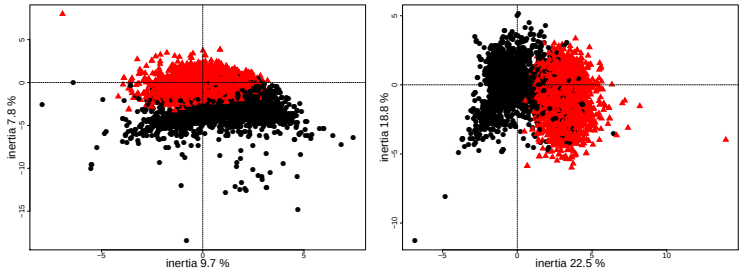
(c) Adj. Rand.: 0.00

Table 2: Adjusted Rand indices and confusion matrices related to: (a) the bi-component heteroscedastic Gaussian copula mixture; (b) the tri-component homoscedastic Gaussian copula mixture; (c) the four-component locally independent mixture.

Visualization Figure 1 displays the individuals in a PCA map of both classes estimated by the bi-component free mixture model of Gaussian copulas. According to these scatter-plots, classes are well-separated.

Interpretation of the best model The following interpretation is based on the margin parameters and on the intra-class correlation matrices summarized in Figure 2. The majority class ($\pi_1 = 0.59$) is principally composed by white wines. This class is characterized by lower rates of acidity, pH, chlorides and sulphites than them of the minority class ($\pi_2 = 0.41$) which is principally composed by red wines. The majority class has larger values for both sulfur dioxide measures and the alcoholic rate. Note than the wine quality of both classes is similar ($\beta_{1\text{quality}} = 5.96$ and $\beta_{2\text{quality}} = 5.58$). The majority class is characterized by a strong correlation between both sulfur measures opposite to a strong correlation between the density and acidity measures. The minority class underlines that the wine quality is dependent with a larger alcoholic rate and small values for the chlorides and acidity measures.

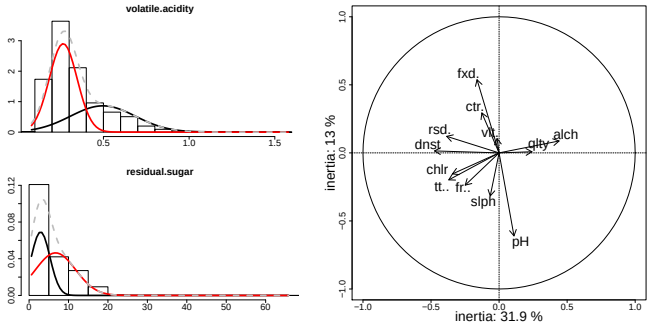
Conclusion On this data set, the mixture model of Gaussian copulas overcomes the locally independent model (reduction of the number of classes, better values of the information criteria,



(a) First PCA of the class 1, map(3,5)

(b) First PCA of the class 2, map(1,2)

Figure 1: Visualization of the partition by the bi-component heteroscedastic mixture model of Gaussian copulas (Class 1 is drawn by black circles and Class 2 by red triangles).



(a) $\pi_k p(x^j|\boldsymbol{\beta}_{kj}, z = k)$

(b) Variables in the first correlation circle given by $\mathbf{\Gamma}_2$

Figure 2: Summary of the bi-component heteroscedastic mixture model of Gaussian copula. Class 1 is drawn in black and Class 2 in red. (fixed acidity: fxd., volatile acidity: vlt., citric acidity: ctr., residual sugar: rsd., chlorides: chlr., free sulfur dioxide: fr., total density dioxide: tt., density: dnst., pH, sulphates: slph., alcohol: alch., quality: qlty.).

estimated partition closest to the true one). Based on the individual scatter-plots in the model PCA, the estimated classes are relevant since they are well-separated. Finally, the estimation of the intra-class dependencies helps the interpretation since it underlines the link between the wine quality of the minority class and its physiochemical properties.

5 Conclusion and future extensions

The proposed model uses the properties of copulas: independent choice of the margin distributions and of the dependency relations. Thus, the mixture model of Gaussian copulas allows to fix classical margins belonging to the exponential family for the component margin distributions and takes into account the intra-class dependencies. An approach based on a PCA per class of the Gaussian latent variable allows to summarize the main intra-class dependencies and to visualize the data by using the model parameters. The application points out that this model

is sufficiently flexible to efficiently fit data and that it can reduces the biases of the locally independent model (for instance the reduction of the number of classes). The number of parameters increases with the number of classes and variables especially because of the correlation matrices of the Gaussian copulas. To avoid this drawback, we propose an homoscedastic version of the model assuming the equality between the correlation matrices. This model may better fit the data than the heteroscedastic Gaussian mixture models.

Bibliography

[1] J.D. Banfield and A.E. Raftery. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, pages 803–821, 1993.

[2] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, 2000.

[3] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553, 2009.

[4] B.S. Everitt. A finite mixture model for the clustering of mixed-mode data. *Statistics & Probability Letters*, 6(5):305–309, 1988.

[5] C. Gouget. *Utilisation des modèles de mélange pour la classification automatique de données ordinales*. PhD thesis, Université de Technologie de Compiègne, 2006.

[6] D.J. Hand and K. Yu. Idiot's Bayes - Not So Stupid after All? *International Statistical Review*, 69(3):385–398, 2001.

[7] P.D. Hoff. Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics*, pages 265–283, 2007.

[8] L. Hunt and M. Jorgensen. Clustering mixed data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4):352–361, 2011.

[9] H. Joe. *Multivariate models and multivariate dependence concepts*, volume 73. CRC Press, 1997.

[10] C.A.J. Klaassen and J.A. Wellner. Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 3(1):55–77, 1997.

[11] D.D. Lewis. Naive (Bayes) at forty: The independence assumption in information retrieval. In *Machine learning: ECML-98*, pages 4–15. Springer, 1998.

[12] G.J. McLachlan and D. Peel. *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, 2000.

[13] U. Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460, 1979.

[14] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.