Computational Statistics and Data Analysis 71 (2014) 196–210

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Parsimonious skew mixture models for model-based clustering and classification

Irene Vrbik, Paul D. McNicholas *

Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

ARTICLE INFO

Article history:
Received 1 July 2012
Received in revised form 4 July 2013
Accepted 4 July 2013
Available online 11 July 2013

Keywords:
Eigenvalue decomposition
EM algorithm
GPCM
MCLUST
Mixture models
Model-based clustering
Skew-normal distribution
Skew-*t* distribution

ABSTRACT

Robust mixture modeling approaches using skewed distributions have recently been explored to accommodate asymmetric data. Parsimonious skew-*t* and skew-normal analogues of the GPCM family that employ an eigenvalue decomposition of a scale matrix are introduced. The methods are compared to existing models in both unsupervised and semi-supervised classification frameworks. Parameter estimation is carried out using the expectation–maximization algorithm and models are selected using the Bayesian information criterion. The efficacy of these extensions is illustrated on simulated and real data sets.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The objective of cluster analysis is to organize data into groups wherein the similarity within groups and the dissimilarity between groups are maximized. A 'model-based' approach is one that uses mixture models for clustering. The use of finite mixture models has become increasingly common for clustering and classification, particularly with the use of Gaussian components. The Gaussian model-based clustering likelihood is

$$\mathcal{L}(\boldsymbol{\vartheta} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{j=1}^{n} \sum_{i=1}^{g} \pi_i \phi(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{1}$$

where $\pi_i > 0$, such that $\sum_{i=1}^{g} \pi_i = 1$, are mixing proportions and $\phi(\mathbf{x}_j \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the density of a multivariate Gaussian random variable with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. Model-based classification is a semi-supervised analogue of model-based clustering (cf. Section 6).

Gaussian mixture models have been used for a wide variety of clustering applications, including work by McLachlan and Basford (1988), Bouveyron et al. (2007), McNicholas and Murphy (2008, 2010a,b), and Baek and McLachlan (2010), amongst others. In efforts to accommodate data that exhibit some departure from normality, robust extensions are garnering increased attention. For instance, mixtures of multivariate t-distributions (McLachlan and Peel, 1998; Peel and McLachlan, 2000) have proven effective for dealing with components containing outliers. They have been the basis of a variety of robust

* Corresponding author. Tel.: +1 519 8244120x53136; fax: +1 519 8370221.
E-mail addresses: vrbiki@uoguelph.ca (I. Vrbik), pmcnicho@uoguelph.ca, paul.mcnicholas@uoguelph.ca (P.D. McNicholas).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.07.008

clustering and classification techniques, including work by McLachlan et al. (2007), Greselin and Ingrassia (2010), Andrews and McNicholas (2011a,b), Andrews et al. (2011), Baek and McLachlan (2011), Steane et al. (2012), McNicholas and Subedi (2012) and Morris et al. (in press).

Capturing components that are asymmetric can be tackled using skew-normal distributions (cf. Lin et al., 2007) or other non-elliptically contoured distributions (e.g., Karlis and Santourian, 2009). One example of a non-elliptical distribution is the skew-normal independent distribution, as considered for finite mixture modeling in Cabral et al. (2012). Another interesting alternative is the skew Student-t-normal distribution that has recently been used to model skewed heavy-tailed data (Ho et al., 2012). Although many non-symmetric options are available, we focus on mixtures of skew-normal distributions and mixtures of skew-t distributions herein.

Recently, mixtures of multivariate skew-t distributions have received some attention within the literature. Of course, the skew-normal, t, and Gaussian distributions are all special or limiting cases of the skew-t distribution. This property can be important in clustering applications because we often do not know the most appropriate underlying distribution. Alongside the skewness parameter that accommodates asymmetric data, the degrees of freedom parameter allows for heavy tails, giving less weight to outlying observations in parameter estimation. Compared to work on parsimonious Gaussian and t-mixtures, the literature contains relatively little on parsimonious mixtures of multivariate skew-normal and skew-t distributions. The purpose of this paper is to go some way towards addressing this deficiency.

The remainder of the paper is organized as follows. In Section 2, we introduce skew-t and skew-normal mixture models and briefly discuss parameter estimation. Section 3 presents the construction of parsimonious families of models that are analogues of popular Gaussian approaches. The proposed methods are compared with Gaussian and multivariate t analogues using simulation studies (Section 4) and four benchmark clustering data sets (Section 5). These models are extended further to semi-supervised classification in Section 6, which is followed by concluding remarks (Section 7).

2. Mixtures of skew-t and skew-normal distributions

2.1. A mixture of skew-t distributions

Although the computational tractability of Gaussian mixture models has contributed to their widespread popularity within the literature, their application is not always appropriate. For instance, Kotz and Nadarajah (2004) argue that the multivariate-t distribution provides a more realistic model for real-world data and it has been noted (e.g., Lin et al., 2007) that Gaussian mixture models have a tendency to overfit skewed data. Therefore, it is natural to consider a single distribution, namely the multivariate skew-t distribution, that conflates the robust properties of the t-distribution with a skewness parameter to account for asymmetry. We outline a model-based approach using parsimonious mixtures of multivariate skew-t distributions and we also consider a mixture of skew-normal distributions, which is a limiting case.

There are a number of 'skew-t' distributions within the literature. The version that we adopt, as defined by Pyne et al. (2009), uses a particular stochastic representation of the multivariate skew-t distribution given by Sahu et al. (2003). Adopting this characterization, a random vector $\boldsymbol{Y}$ is said to follow a p-variate skew-t distribution with location vector $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Omega}$, skewness vector $\boldsymbol{\lambda}$, and ν degrees of freedom if, conditional on a random variable $W \sim \text{gamma}(\nu/2, \nu/2)$, it has the representation

$$\boldsymbol{Y} = \boldsymbol{\lambda}|U| + \boldsymbol{X}, \tag{2}$$

where

$$\begin{pmatrix} \boldsymbol{X} \\ U \end{pmatrix} \Bigg| W = w \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\xi} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ 0 & 1 \end{pmatrix} \frac{1}{w} \right).$$

We consider a g-component mixture of p-dimensional skew-t distributions with density given by

$$f(\boldsymbol{y}_j \mid \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i \varkappa(\boldsymbol{y}_j \mid \boldsymbol{\xi}_i, \boldsymbol{\Omega}_i, \boldsymbol{\lambda}_i, \nu_i), \tag{3}$$

where $\varkappa(\boldsymbol{y}_j \mid \boldsymbol{\xi}_i, \boldsymbol{\Omega}_i, \boldsymbol{\lambda}_i, \nu_i)$ is the density of a multivariate skew-t distribution with location vector $\boldsymbol{\xi}_i$, scale matrix $\boldsymbol{\Omega}_i$, skewness parameter $\boldsymbol{\lambda}_i$, and ν_i degrees of freedom, and $\boldsymbol{\Psi}$ contains all model parameters, i.e., $\boldsymbol{\Psi}$ contains the parameters $\{\pi_i, \boldsymbol{\xi}_i, \boldsymbol{\Omega}_i, \boldsymbol{\lambda}_i, \nu_i : i = 1, \ldots, g\}$.

2.2. A mixture of skew-normal distributions

Now consider the skew-normal distribution, which is a limiting case of the skew-t distribution. The version we adopt is defined by Sahu et al. (2003) and proposed for the analysis of flow cytometric data by Pyne et al. (2009). Resembling the above characterization (Section 2.1), a random vector $\boldsymbol{Y}$ is said to follow a p-variate skew-normal distribution with location vector $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Omega}$, and skewness vector $\boldsymbol{\lambda}$ if it has the representation

$$\boldsymbol{Y} = \boldsymbol{\lambda}|U| + \boldsymbol{X}, \tag{4}$$

Download English Version:

https://daneshyari.com/en/article/6870340

Download Persian Version:

https://daneshyari.com/article/6870340

Daneshyari.com