



Outlier Detection via Parsimonious Mixtures of Contaminated Gaussian Distributions

Antonio Punzo, Paul D. McNicholas

(Submitted on 20 May 2013)

For multivariate continuous data, the contaminated Gaussian distribution - having two parameters indicating the proportion of outliers and the degree of contamination - represents a convenient and natural way to model and detect outliers. In this paper, we introduce a mixture model whereby each mixture component is itself a contaminated Gaussian distribution. To introduce parsimony, a family of fourteen mixtures of contaminated Gaussian distributions is developed by applying constraints to eigen-decomposed component covariance matrices. This approach is, amongst other things, an effective alternative to trimmed clustering. Although these models could be used for model-based clustering, classification, and discriminant analysis, we focus on the more general model-based classification framework. An expectation-conditional maximization algorithm is used to find maximum likelihood estimates of the parameters and thereby give classifications for the observations. A simulation study is performed to evaluate the behaviour of the Bayesian information criterion and the integrated completed likelihood in model selection. This novel family of models is applied to artificial and real data in order to illustrate some of its advantages. Amongst them, and in contrast to the trimmed clustering approach, we have: 1) each observation has a posterior probability of belonging to a particular group and, inside each group, of being an outlier or not, 2) the models do not require pre-specification of quantities such as the proportion of observations to trim, 3) the approach can be easily used in high dimensions, and 4) model-based classification is permitted in addition to clustering.

Download:

- [PDF](#)
- [PostScript](#)
- [Other formats](#)

Current browse context:

stat.ME

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1305](#)

Change to browse by:

[stat](#)

[stat.CO](#)

References & Citations

- [NASA ADS](#)

Bookmark [\(what is this?\)](#)



Subjects: **Methodology (stat.ME)**; Computation (stat.CO)