Trimmed Likelihood Fitting of Mixture Models

N. Neykov[1], P. Filzmoser[2] R. Dimova[1], and P. Neytchev[1]

[1] Bulgarian Academy of Sciences, 66 Tsarigradsko chaussee, Sofia 1784, Bulgaria
[2] Vienna University of Technology, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria

Keywords: Mixture of Distributions, Trimmed Likelihood Estimator

Abstract

Finite mixture of distributions have been widely used to model a wide range of heterogeneous data. Details can be found in McLachlan and Peel (2000). In most applications the unknown mixture model parameters are estimated by the classical Maximum Likelihood Estimator (MLE). However, the MLE can be very sensitive to outliers in the data. In this paper we consider an approach based on the Trimmed Likelihood Estimator (TLE) introduced by Neykov and Neytchev (1990) and further developed by Vandev and Neykov (1993), and Müller and Neykov (2003) to estimate mixture of multivariate normals and Generalized Linear Models (GLMs) in a robust way. The TLE is defined over those k observations out of n with the largest MLE fit thus those observations that do not follow the assumed model are trimmed. Because of its combinatorial nature computing the TLE is infeasible for large data sets. To get approximate parameter estimates of the mixture components the FAST-TLE algorithm developed by Neykov and Müller (2002) is adapted. The FAST-TLE algorithm essentially consists of carrying out finitely many times a two-step MLE procedure based on a trial step followed by a refinement step therefore any program for ML fitting of mixture models and model-based cluster analysis can be used as a computational engine. For this purposes the FlexMix program of Leisch (2004) is used. Examples of real and artificial data in the context of clusterwise linear, binary, binomial and Poisson regressions, and multivariate normals are used to illustrate the superiority of the TLE approach versus the MLE. The breakdown point of the TLE in the above model frameworks will be discussed as well.

References

F. Leisch (2004). *FlexMix Manual:* http://cran.R-project.org/doc/packages/flexmix.pdf.

G. McLachlan and D. Peel (2000). *Finite mixture models.* Wiley & Sons, New York.

C. Müller and N. Neykov (2003). Breakdown Points of the Trimmed Likelihood and Related Estimators in GLMs. *J. Statist. Plann. Inference.*, 116, 503–519.

N. Neykov, P. Filzmoser, R. Dimova and P. Neytchev (2004). Mixture of GLMs and the Trimmed Likelihood Methodology. In: J. Antoch, editor, *Proc. in Computational Statistics*, pp. 1585–1592, Physica Verlag Berlin, Germany.

N. Neykov and C. Müller (2002). Breakdown Point and Computation of Trimmed Likelihood Estimators in GLMs. In: R. Dutter et al., editors, *Developments in robust statistics*, pp. 277–286, Physica Verlag, Heidelberg.

N. Neykov and P. Neytchev (1990). A robust alternative of the MLE. In: Short communications of COMPSTAT'90, Dubrovnik, pp. 99–100.

D. Vandev and N. Neykov (1993). Robust Maximum Likelihood in the Gaussian Case. In: S. Morgenthaler et al., editors, *New Directions in Data Analysis and Robustness*, pp. 259–264, Birkhäuser Verlag Basel, Switzerland.