TO GET THE BEST, TAME THE BEAST: ROBUST ML ESTIMATION FOR MIXTURE MODELS

Francesca Greselin [1]

[1] Department of Statistics and Quantitative Methods, University of Milano Bicocca, (e-mail: francesca.greselin@unimib.it)

ABSTRACT: This paper presents a brief review of constrained maximization of the likelihood, in combination with data-driven trimming, as a powerful technique for achieving robust classification and clustering in mixture models. Trimming is the simpler way to achieve robustness, being also highly intuitive. Originally developed for Gaussian components, this methodology has been successfully extended to various scenarios, including parsimonious mixtures, mixtures of factor analyzers, mixtures of regression and cluster-weighted mixtures, as well as to mixtures of skew and functional data. By effectively taming the complexities associated with parameter estimation, this approach yields an estimator which exists and is strongly consistent to the corresponding solution of the population optimum under widely general conditions.

KEYWORDS: Model-based classification, robustness, trimming, constrained estimation, outliers.

1 Introduction

Mixture models offer a highly flexible approach for statistical modeling of diverse random phenomena, especially when we posit that the observations arise from unobserved groups within the population. However, estimating Gaussian (and related) mixture models using the Maximum Likelihood (ML) approach introduces two significant challenges: *i)* the unboundedness of the likelihood function, that sets the ML as a mathematically ill-posed problem, and *ii)* the presence of contaminating data (background noise, pointwise contamination, unexpected minority patterns, etc.) that could severely affect the model fitting.

To *tame* the likelihood, researchers have adopted two essential techniques for parameter estimation in mixture models: eigenvalue constraints and trimming. The foundation of this methodology can be traced back to García-Escudero *et al.*, 2008, which has since evolved into a paradigm for robust model-based classification and clustering. The approach is well-regarded for its desirable theoretical properties and the availability of feasibile EM algorithms for its implementation. Constrained estimation prevents convergence towards degenerate solutions, and mitigates the occurrence of non-interesting (spurious) local maximizers associated with complex likelihood surfaces. From Hathaway's seminal paper (Hathaway, 1985), these approaches, when coupled with impartial trimming, find applications in the context of robust statistical methods. Impartial trimming consists in excluding a small percentage of the less plausible observations, during the EM iterations, from contributing to model estimation. So doing, it protects the inferential results from the harmful effects of outliers.

Within the realm of this research stream, notable contributions include robust mixtures of factor analyzers (García-Escudero *et al.*, 2017) and the robust cluster-weighted model (García-Escudero *et al.*, 2016b). The introduction of their fuzzy versions (García-Escudero *et al.*, 2018b) revealed an intriguing interplay between the fuzzifier parameter and the scale. Additionally, advancements have been made in the treatment of skew components (García-Escudero *et al.*, 2016a). In the context of the semisupervised setting, when label noise interferes with the learning process, and whenever variable selection could be beneficial, the development of a specific robust approach is needed (Cappozzo *et al.*, 2020b, Cappozzo *et al.*, 2021).

However, in all such models, hyperparameter tuning remains an essential part of the inferential process, and ongoing research is focused on determining optimal settings for critical parameters such as the percentage of trimming, the number of components in the mixture, and the value for the eigenvalue constraint (Riani *et al.*, 2019). To support practitioners in this delicate task, researchers have proposed graphical and computational tools based on the combination of two exploratory steps (Cappozzo *et al.*, 2023).

Indeed, the literature presents several alternative methodologies for achieving robust model-based classification. Some of these approaches involve substituting Gaussian components with other elliptical distributions that possess heavier tails, such as the Student t (e.g., Greselin & Ingrassia, 2010) or the contaminated normal distribution (Punzo & McNicholas, 2016). These approaches withstand the presence of mild outliers. One key concept used to assess the robustness of an estimate in the presence of outliers is the breakdown point (Hampel, 1971). Its finite sample version is the maximum fraction of outliers which a given sample may contain without spoiling the estimate completely (Donoho, 1982). Among the models with good breakdown properties, we may mention a method for cluster detection and clustering with random start forward searches (Atkinson *et al.*, 2018), the optimally tuned robust improper maximum likelihood estimator, which uses an improper constant density for modeling outliers and noise (Coretto & Hennig, 2017), and the weighted likelihood approach, aimed at downweighting outliers (Greco & Agostinelli, 2020). Here the weights are based on Pearson residuals stemming from robust Mahalanobis-type distances.

To conclude, the reviewed models deliver more reliable and stable estimation in the presence of outliers and noisy data, significantly enhancing model performance and facilitating more accurate statistical inference. The field of robust model-based clustering and classification is continually progressing, built upon solid results and theoretical foundations, while still presenting numerous intriguing challenges that await exploration. Exciting possibilities lie ahead, and the best is yet to come for researchers in this evolving domain.

References

ATKINSON, A.C., RIANI, M., & CERIOLI, A. 2018. Cluster detection and clustering with random start forward searches. *Journal of Applied Statistics*, **45**(5), 777 – 798.

CAPPOZZO, A., GRESELIN, F., & MURPHY, T.B. 2020a. A robust approach to model-based classification based on trimming and constraints: Semi-supervised learning in presence of outliers and label noise. *Advances in Data Analysis and Classification*, **14**(2), 327–354.

CAPPOZZO, A., GRESELIN, F., & MURPHY, T.B. 2020b. Anomaly and Novelty detection for robust semi-supervised learning. *Statistics and Computing*, **30**(5), 1545–1571.

CAPPOZZO, A., GRESELIN, F., & MURPHY, T.B. 2021. Robust variable selection for model-based learning in presence of adulteration. *Computational Statistics & Data Analysis*, **158**, 107–186.

CAPPOZZO, A., GARCÍA-ESCUDERO, L.A., GRESELIN, F., & MAYO-ISCAR, A. 2023. Graphical and computational tools to guide parameter choice for the cluster weighted robust model. *Journal of Computational and Graphical Statistics*, 1–20.

CORETTO, P., & HENNIG, C. 2017. Consistency, breakdown robustness, and algorithms for robust improper maximum likelihood clustering. *Journal of Machine Learning Research*, **18**(142), 1–39.

DONOHO, D.L. 1982. Breakdown properties of multivariate location estimators. *Ph.D. qualifying paper, Harvard University*.

GARCÍA-ESCUDERO, L. A., GORDALIZA, A., GRESELIN, F., INGRASSIA, S., & MAYO-ÍSCAR, A. 2017. Robust estimation of mixtures of regressions with random covariates, via trimming and constraints. *Statistics and Computing*, **27**(2), 377–402.

GARCÍA-ESCUDERO, L.A., GORDALIZA, A., MATRÁN, C., & MAYO-ISCAR, A. 2008. A general trimming approach to robust cluster analysis. *Annals of Statistics*, **36**(3), 1324–1345.

GARCÍA-ESCUDERO, L.A., GRESELIN, F, MC LACHLAN, G., & MAYO-ISCAR, A. 2016a. Robust estimation of mixtures of Skew Normal Distributions. *In: Proceedings of the 48th Scientific Meeting of the Italian Statistical Society-Salerno (Italy), June 8-10, 2016*.

GARCÍA-ESCUDERO, L.A., GORDALIZA, A., GRESELIN, F., INGRASSIA, S., & MAYO-ISCAR, A. 2016b. The joint role of trimming and constraints in robust estimation for mixtures of Gaussian factor analyzers. *Computational Statistics & Data Analysis*, **99**, 131–147.

GARCÍA-ESCUDERO, L.A., GORDALIZA, A., GRESELIN, F., INGRASSIA, S., & MAYO-ISCAR, A. 2018a. Eigenvalues and constraints in mixture modeling: geometric and computational issues. *Advances in Data Analysis and Classification*, **12**(2), 203–233.

GARCÍA-ESCUDERO, L.A., GRESELIN, F., & MAYO-ISCAR, A. 2018b. Robust, fuzzy, and parsimonious clustering, based on mixtures of factor analyzers. *International Journal of Approximate Reasoning*, **94**, 60–75.

GRECO, L., & AGOSTINELLI, C. 2020. Weighted likelihood mixture modeling and model-based clustering. *Statistics and Computing*, **30**(2), 255–277.

GRESELIN, F., & INGRASSIA, S. 2010. Constrained monotone EM algorithms for mixtures of multivariate *t* distributions. *Statistics and Computing*, **20**(1), 9–22.

GRESELIN, F., & INGRASSIA, S. 2015. Maximum likelihood estimation in constrained parameter spaces for mixtures of factor analyzers. *Statistics and Computing*, **25**(2), 215–226.

HAMPEL, F.R. 1971. A general qualitative definition of robustness. *The annals of mathematical statistics*, **42**(6), 1887–1896.

HATHAWAY, R.J. 1985. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, **13**(2), 795–800.

PUNZO, A., & MCNICHOLAS, P.D. 2016. Parsimonious mixtures of multivariate contaminated normal distributions. *Biometrical Journal*, **58**(6), 1506–1537.

RIANI, M., ATKINSON, A.C., CERIOLI, A., & CORBELLINI, A. 2019. Efficient robust methods via monitoring for clustering and multivariate data analysis. *Pattern Recognition*, **88**, 246 – 260.