

Robust variable selection for finite mixture regression models

Qingguo Tang¹ · R. J. Karunamuni²

Received: 18 May 2016 / Revised: 29 September 2016 / Published online: 25 February 2017
© The Institute of Statistical Mathematics, Tokyo 2017

Abstract Finite mixture regression (FMR) models are frequently used in statistical modeling, often with many covariates with low significance. Variable selection techniques can be employed to identify the covariates with little influence on the response. The problem of variable selection in FMR models is studied here. Penalized likelihood-based approaches are sensitive to data contamination, and their efficiency may be significantly reduced when the model is slightly misspecified. We propose a new robust variable selection procedure for FMR models. The proposed method is based on minimum-distance techniques, which seem to have some automatic robustness to model misspecification. We show that the proposed estimator has the variable selection consistency and oracle property. The finite-sample breakdown point of the estimator is established to demonstrate its robustness. We examine small-sample and robustness properties of the estimator using a Monte Carlo study. We also analyze a real data set.

Keywords Finite mixture regression models · Variable selection · Minimum-distance methods

✉ R. J. Karunamuni
r.j.karunamuni@ualberta.ca

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

² Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada