

Integrating multi-source block-wise missing data in model selection

Fei Xue

University of Illinois
E-mail: xuefei012@gmail.com

Abstract: For multi-source data, blocks of variable information from certain sources are likely missing. Existing methods for handling missing data do not take structures of block-wise missing data into consideration. In this paper, we propose a Multiple Block-wise Imputation (MBI) approach, which incorporates imputations based on both complete and incomplete observations. Specifically, for a given missing pattern group, the imputations in MBI incorporate more samples from groups with fewer observed variables in addition to the group with complete observations. We propose to construct estimating equations based on all available information, and optimally integrate informative estimating functions to achieve efficient estimators. We show that the proposed method has estimation and model selection consistency under both fixed-dimensional and high-dimensional settings. Moreover, the proposed estimator is asymptotically more efficient than the estimator based on a single imputation from complete observations only. In addition, the proposed method is not restricted to missing completely at random. Numerical studies and ADNI data application confirm that the proposed method outperforms existing variable selection methods under various missing mechanisms.