

Empirical Bayes Variable Selection Using Iterative Conditional Modes/Medians

Vitara Pungpapong

Department of Statistics, Chulalongkorn University, Bangkok, Thailand vitara@cbs.chula.ac.th

Min Zhang

Department of Statistics, Purdue University, West Lafayette, IN, USA minzhang@purdue.edu
Center of Bioinformatics, Capital Medical University, Beijing, China

Dabao Zhang*

Department of Statistics, Purdue University, West Lafayette, IN, USA zhangdb@purdue.edu
Center of Bioinformatics, Capital Medical University, Beijing, China

We proposed an empirical Bayes variable selection for high-dimensional data, which incorporates sparsity or more complicated *a priori* information. The empirical Bayes thresholding is extended beyond independent data for its property that it bears no tuning parameters. An iterated conditional modes/medians (ICM/M) algorithm is proposed to implement the empirical Bayes variable selection by iteratively minimizing a conditional loss function. The iterative conditional modes are employed to obtain data-driven estimates of hyperparameters, and the iterative conditional medians are used to estimate the model coefficients and therefore enable the selection of massive variables. The ICM/M algorithm is computationally fast. Empirical studies suggest very competitive performance of the proposed method, even in the simple case of selecting massive regression predictors.

Key Words: High dimensional data; iterated conditional modes/medians; linear regression; sparsity; structured variables