

## Recursive bootstrap $L_1$ -type regularized regression modeling based on parametric statistical test

Heewon Park\*, Seiya Imoto and Satoru Miyano  
Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan  
hwpark@ims.u-tokyo.ac.jp

### Abstract

The  $L_1$ -type regularization approaches have been widely used for uncovering cancer driver gene based on various genome-scale information. Although the existing  $L_1$ -type regularization methods have been widely used to various fields of research, there are several drawbacks as a tool for feature selection in high dimensional data analysis: limitation of subset size, erroneous estimation result, multicollinearity problem and time consuming procedures. We propose a novel statistical strategy, called a Recursive Random Lasso (RRLasso) for high dimensional data analysis in line with a random lasso. In order to time effective analysis, we consider recursive bootstrap procedure based on random forest method. Furthermore, we introduce a parametric statistical test for variable selection based on bootstrap regression modeling results. We can see through Monte Carlo simulations that the proposed RRLasso not only provides time effective performances but also performs well for high dimensional data analysis.

**Keywords:** bootstrap method;  $L_1$ -type regularization; parametric statistical test; random forest method.