Random Lasso

by
Sijian Wang
University of Wisconsin, Madison
600 Highland Ave.
Madison, WI 53705
swang@biostat.wisc.edu

Abstract

We propose a computationally intensive method, the random lasso method, for variable selection in linear models. The method consists of two major steps. In Step 1, the lasso method is applied to each of the bootstrap samples for a set of randomly selected covariates. A measure of importance is yielded from this step for each covariate. In Step 2, a similar procedure to the first step is implemented with the exception that for each bootstrap sample, a subset of covariates is randomly selected with unequal selection probabilities determined by covariates' importance measures. Adaptive lasso may be used in the second step with weights determined by the importance measures. The final set of covariates and their coefficients are determined by averaging bootstrap results obtained from Step 2. The proposed method alleviates some of the limitations of lasso and related methods: it tends to remove highly correlated variables altogether or select them all, and maintains maximal fle! xibility in estimating their coefficients, particularly with different signs; and the number of selected variables is no longer limited by the sample size. We illustrate the proposed method by extensive simulation studies. The proposed method is also applied to a Glioblastoma microarray data analysis.

This is a joint work with Bin Nan, Saharon Rosset and Ji Zhu.