

“Group Selection in High-dimensional Regression”

Fengrong Wei

Department of Mathematics
The University of Iowa

Abstract

We consider the problem of selecting grouped variables for accurate prediction in regression. Such a problem arises naturally in many practical situations with microarray data analysis as one well-known example. Group Lasso is an attractive method for such kind regression problem, since it respects the grouping structure in the data.

We study the selection and estimation properties of the group Lasso in high-dimensional settings when the number of groups exceeds the sample size. Under appropriate conditions, we show that the group Lasso selects a model of the right order of dimensionality, selects all variables whose coefficient norms are greater than certain threshold level, and is estimation consistent. However, the group Lasso is in general not selection consistent and tends to select variables that are not significant in the model. In order to improve the selection results, we apply the adaptive group Lasso. We show that under suitable conditions, the adaptive group Lasso has oracle selection property, in the sense that it can correctly select important variables with probability converging to one. In contrast, group Lasso do not possess such oracle property.