

Correlated Variable Selection in High-dimensional Linear Models using Dual Polytope Projection

Niharika Gauraha and Swapan K. Parui

Indian Statistical Institute, India

We consider the case of high dimensional linear models ($p \gg n$) with strong empirical correlation among variables. The Lasso is a widely used regularized regression method for variable selection, but it tends to select a single variable from a group of strongly correlated variables even if many or all of these variables are important. In many situations, it is desirable to identify all the relevant correlated variables, examples include micro-array analysis and genome-wide association studies. We propose to use Dual Polytope Projections (DPP) rule, for selecting the relevant correlated variables which are not selected by the Lasso.

We consider the usual linear model setup, that is given as $\mathbf{Y} = \mathbf{X}\beta + \epsilon$. Let $\lambda \geq 0$ be a regularization parameter. Then the Lasso estimator(see [1]) is defined as: $\hat{\beta}(\lambda) = \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$. Let $\lambda_{max} = \max_{1 \leq j \leq p} |\mathbf{X}_j^T \mathbf{Y}|$, then for all $\lambda \in [\lambda_{max}, \infty)$, we have $\hat{\beta}(\lambda) = 0$. It has been shown that the screening methods based on DPP rule are highly effective in reducing the dimensionality by discarding the irrelevant variables (see [2]). Suppose we want to compute Lasso solution for a $\lambda \in (0, \lambda_{max})$, the (global strong) DPP rule discards the j^{th} variable whenever $|\mathbf{X}_j^T \mathbf{Y}| < 2\lambda - \lambda_{max}$ (variables having smaller inner products with the response).

Exploiting the above property, we propose a two-stage procedure for variable selection. At the first stage, we perform Lasso using cross-validation and we choose the regularization parameter λ_{Lasso} , that optimizes the prediction. At the second stage, we select all the variables for which $|\mathbf{X}_j^T \mathbf{Y}| \geq 2\lambda_{Lasso} - \lambda_{max}$. Though, the Lasso solution at λ_{Lasso} does not include all the relevant correlated variables, but these correlated variables have the similar magnitude for their inner products with the response. Hence, all the relevant correlated predictors also get selected at the second stage.

Keywords: Correlated Variable Selection, Lasso, Dual Polytope Projection, High-dimensional Data Analysis

References

- [1] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *J. R. Statist. Soc* 58 (1996), 267–288.
- [2] Jie Wang et al. "Lasso Screening Rules via Dual Polytope Projection". In: *NIPS* (2013).