Adaptive Generalized Lasso for High-Dimensional Linear Regression Model

Tokyo Institute of Technology Shota Katayama

Modern data analysts often suffer from the high dimensionality, where the data have larger dimension p than the sample size n. One of the most popular approach in the context of linear regression would be the lasso proposed by Tibshirani (1996). Recently, a large number of studies have supported its utility under the assumption that only a subset of covariates is relevant to responses. Though this sparsity assumption is natural and appropriate for high dimensional data, it is not enough for capturing more complex structures among covariates. In many cases, data analysts have a prior knowledge about covariates such as the order and (dis)similarity. To utilize such a knowledge, the adaptive generalized lasso (AGL, hereafter) given as

$$\operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2n}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{\Lambda}_1\boldsymbol{\beta}\|_1 + \|\boldsymbol{\Lambda}_2\boldsymbol{C}\boldsymbol{\beta}\|_1, \tag{1}$$

is studied in this talk. This is an adaptive version of generalized lasso proposed by Tibshirani and Taylor (2012). The matrix $\boldsymbol{C} \in \mathbb{R}^{q\times p}$ is known and enables to handle complex structures among covariates. For instance, when $(\boldsymbol{C})_{i,i} = 1$, $(\boldsymbol{C})_{i,i+1} = -1$ for $i = 1,\ldots,p-1$ and $(\boldsymbol{C})_{j,k} = 0$ otherwise, an order of covariates is utilized. For a given graph with edges $E_1,\ldots,E_q$, when $(\boldsymbol{C})_{i,j} = 1$, $(\boldsymbol{C})_{i,k} = -1$ for $E_i = (j,k)$ and $(\boldsymbol{C})_{i,\ell} = 0$ otherwise, then similarities among covariates on the graph are utilized. The matrices $\boldsymbol{\Lambda}_1 = \lambda_1\mathrm{diag}(w_{11},\ldots,w_{1p})$ and $\boldsymbol{\Lambda}_2 = \lambda_2\mathrm{diag}(w_{21},\ldots,w_{2q})$ denote weights for $\boldsymbol{\beta}$ and $\boldsymbol{C}\boldsymbol{\beta}$, respectively. These weights are calculated via the non-adaptive generalized lasso as

$$\operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2n}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \tau\|\boldsymbol{D}\boldsymbol{\beta}\|_1, \quad \boldsymbol{D} = \begin{pmatrix}\boldsymbol{I}_p\\ \boldsymbol{C}\end{pmatrix}. \tag{2}$$

An important question is how AGL given in (1) with some weights via (2) performs under the high dimensionality. This talk shows that AGL has the consistency and support recovery if the matrix $\boldsymbol{C}$ satisfies some conditions. Some examples of $\boldsymbol{C}$ meeting them are also introduced.

References

[1] R. Tibshirani. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society B*, **58**, 267–288.

[2] R. J. Tibshirani and J. Taylor. (2011). The solution path of the generalized lasso. *The Annals of Statistics*, **39**, 1335–1371.