Robust high-dimensional regression with algorithmic convergence and support recovery

Shota Katayama, Tokyo Institute of Technology
Hironori Fujisawa, The Institute of Statistical Mathematics

Linear regression with a large number of covariates is a general and fundamental problem in recent data analysis. When outliers exist in responses, however, a standard method such as the Lasso does not perform well. In this talk, we consider robust alternatives based on the linear regression model with outliers

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \sqrt{n}\boldsymbol{\gamma}^* + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ is an n dimensional response vector, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ is an $n \times p$ covariate matrix, $\boldsymbol{\beta}^*$ is a p dimensional unknown coefficient vector, $\boldsymbol{\gamma}^*$ is an n dimensional unknown vector whose nonzero elements correspond to outliers and $\boldsymbol{\varepsilon}$ is an n dimensional random error vector.

Let $L(\boldsymbol{\beta}; \boldsymbol{\gamma}) = \frac{1}{2n}\|(\boldsymbol{y} - \sqrt{n}\boldsymbol{\gamma}) - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_\beta \sum_{j=1}^p w_{\beta,j}|\beta_j|$ and $[\sqrt{n}h(\boldsymbol{\beta})]_i = \Theta(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}; \lambda_\gamma w_{\gamma,i})$ for $i = 1, \ldots, n$, where $\lambda_\beta > 0$ and $\lambda_\gamma > 0$ are tuning parameters, $w_{\beta,j} \geq 0$ and $w_{\gamma,i} \geq 0$ are weights and $\Theta(z; \lambda)$ is a thresholding function. For instance, the soft thresholding is given by $\Theta(z; \lambda) = \mathrm{sgn}(z)\max(|z| - \lambda, 0)$ and the hard thresholding is given by $\Theta(z; \lambda) = zI(|z| > \lambda)$ where $\mathrm{sgn}(\cdot)$ denotes the sign function and $I(\cdot)$ denotes the indicator function. To obtain an estimate of $\boldsymbol{\beta}^*$, we introduce the following algorithm.

Algorithm 1

Step 1. Initialize $k \leftarrow 0$, $\boldsymbol{\beta}^k \leftarrow \boldsymbol{\beta}^{init}$.

Step 2. Update $k \leftarrow k + 1$,

$$\boldsymbol{\beta}^k \leftarrow \mathrm{argmin}_{\boldsymbol{\beta}} L\big(\boldsymbol{\beta}; h(\boldsymbol{\beta}^{k-1})\big).$$

Step 3. Repeat Step 2 until it converges.

Under suitable conditions for the covariates, the random error vector, the thresholding function and the weights, it is shown that

$$\mathbb{P}\left(\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^*\|_2 \leq C\sqrt{\frac{s^* \log p}{n}}\right) \to 1, \quad \mathbb{P}\big\{\mathrm{supp}(\boldsymbol{\beta}^k) = \mathrm{supp}(\boldsymbol{\beta}^*)\big\} \to 1,$$

after some iterations, where $s^* = \mathrm{supp}(\boldsymbol{\beta}^*)$ and $\mathrm{supp}(\boldsymbol{\beta}^*)$ denotes the support of $\boldsymbol{\beta}^*$. The first property shows that the convergence rate of $\boldsymbol{\beta}^k$ is equivalent to the rate of the standard Lasso excluding the true outlier $\sqrt{n}\boldsymbol{\gamma}^*$ from the model in advance. The second property shows that the estimate $\boldsymbol{\beta}^k$ can recover the true support after some iterations. The details can be found in Katayama and Fujisawa (2015, arXiv:1505.05257).