JMLR: Workshop and Conference Proceedings vol 49:1–2, 2016

Aggregation of supports along the Lasso path

Pierre C. Bellec PIERRE.BELLEC@ENSAE.FR
CREST-ENSAE, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France

Abstract

[1] In linear regression with fixed design, we propose two procedures that aggregate a data-driven collection of supports. The collection is a subset of the 2^p possible supports and both its cardinality and its elements can depend on the data. The procedures satisfy oracle inequalities with no assumption on the design matrix. Then we use these procedures to aggregate the supports that appear on the regularization path of the Lasso in order to construct an estimator that mimics the best Lasso estimator. If the restricted eigenvalue condition on the design matrix is satisfied, then this estimator achieves optimal prediction bounds. Finally, we discuss the computational cost of these procedures.

Let n, p be two positive integers. We consider the mean estimation problem

$$Y_i = \mu_i + \xi_i, \qquad i = 1, ..., n,$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^T \in \mathbf{R}^n$ is unknown, $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^T$ is a subgaussian vector, that is,

$$\mathbb{E}[\exp(\boldsymbol{v}^T\boldsymbol{\xi})] \leq \exp\frac{\sigma^2|\boldsymbol{v}|_2^2}{2} \qquad \text{for all } \boldsymbol{v} \in \mathbf{R}^n,$$

where $\sigma > 0$ is the noise level and $|\cdot|_2$ is the Euclidean norm in $\mathbf{R}^n$. We only observe $\mathbf{y} = (Y_1, ..., Y_n)^T$ and wish to estimate $\boldsymbol{\mu}$. A design matrix $\mathbb{X}$ of size $n \times p$ is given and p may be larger than n. We do not require that the model is well-specified, i.e., that there exists $\boldsymbol{\beta}^* \in \mathbf{R}^p$ such that $\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta}^*$. Our goal is to find an estimator $\hat{\boldsymbol{\mu}}$ such that the prediction loss $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2$ is small, where $\|\cdot\|^2$ is the empirical loss defined by

$$\|\boldsymbol{u}\|^2 = \frac{1}{n}|\boldsymbol{u}|_2^2 = \frac{1}{n}\sum_{i=1}^n u_i^2, \qquad \boldsymbol{u} = (u_1, ..., u_n)^T \in \mathbf{R}^n.$$

In a high-dimensional setting where $p > n$, the Lasso is known to achieve good prediction performance. For any tuning parameter $\lambda > 0$, define the Lasso estimate $\hat{\boldsymbol{\beta}}_\lambda^{\mathrm{L}}$ as any solution of the convex minimization problem

$$\hat{\boldsymbol{\beta}}_\lambda^{\mathrm{L}} \in \underset{\boldsymbol{\beta}\in\mathbf{R}^p}{\operatorname{argmin}} \frac{1}{2n}|\mathbf{y} - \mathbb{X}\boldsymbol{\beta}|_2^2 + \lambda\,|\boldsymbol{\beta}|_1\,,$$

where $|\boldsymbol{\beta}|_1 = \sum_{j=1}^n |\beta_j|$ is the ℓ_1-norm. The paper studies an estimator $\hat{\boldsymbol{\mu}}$ that is constructed by aggregation or selection of the Lasso estimators that appear on the regularization path of the Lasso. This estimator $\hat{\boldsymbol{\mu}}$ mimics the best Lasso estimator, in the sense that with probability at least $1 - 1/n$ we have

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \leq C \min_{\lambda>0}\left(\|\mathbb{X}\hat{\boldsymbol{\beta}}_\lambda^{\mathrm{L}} - \boldsymbol{\mu}\|^2 + \frac{\sigma^2(1 + |\hat{\boldsymbol{\beta}}_\lambda^{\mathrm{L}}|_0)}{n}\log\left(\frac{ep}{1 + |\hat{\boldsymbol{\beta}}_\lambda^{\mathrm{L}}|_0}\right)\right),$$

1. Extended abstract. Full version appears as (Bellec, 2016, v2).

where $C > 0$ is a numerical constant and $|\boldsymbol{\beta}|_0$ is the number of nonzero coefficients of any $\boldsymbol{\beta} \in \mathbf{R}^p$.

Keywords: Linear regression, prediction loss, aggregation, Lasso path, oracle inequalities.

References

Pierre C. Bellec. Aggregation of supports along the lasso path. *arXiv:1602.03427*, 2016. URL http://arxiv.org/abs/1602.03427.