Computational Statistics and Data Analysis 142 (2020) 106819

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized ℓ_1-penalized quantile regression with linear constraints

Yongxin Liu [a], Peng Zeng [b,*], Lu Lin [c,d]

[a] Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, China
[b] Department of Mathematics and Statistics, Auburn University, Auburn, United States
[c] School of Statistics, Shandong Technology and Business University, Yantai, China
[d] School of Statistics, Qufu Normal University, Qufu, China

ARTICLE INFO

Article history:
Received 17 March 2018
Received in revised form 19 July 2019
Accepted 20 July 2019
Available online 28 July 2019

Keywords:
Degrees of freedom
Generalized lasso
KKT conditions
Linear programming
Quantile regression

ABSTRACT

In many application areas, prior subject matter knowledge can be formulated as constraints on parameters in order to get a more accurate fit. A generalized ℓ_1-penalized quantile regression with linear constraints on parameters is considered, including either linear inequality or equality constraints or both. It allows a general form of penalization, including the usual lasso, the fused lasso and the adaptive lasso as special cases. The KKT conditions of the optimization problem are derived and the whole solution path is computed as a function of the tuning parameter. A formula for the number of degrees of freedom is derived, which is used to construct model selection criteria for selecting optimal tuning parameters. Finally, several simulation studies and two real data examples are presented to illustrate the proposed method.

1. Introduction

Different from classical regression methods that mainly focus on estimating conditional mean functions, quantile regression is intended for estimating conditional quantile functions and is able to provide a comprehensive picture on how a response depends on predictors. Consider a sample $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ is a p-dimensional vector of predictors and y_i is the response. Quantile regression estimates an intercept $\alpha \in \mathbb{R}$ and a vector of coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$ by solving the following optimization problem,

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \boldsymbol{\beta}^T \boldsymbol{x}_i),$$

where $\tau \in (0, 1)$, ρ_τ is the check function defined by

$$\rho_\tau(u) = \tau u I(u \geq 0) - (1-\tau) u I(u < 0),$$

and $I(\cdot)$ is the indicator function. Refer to Koenker (2005) for a comprehensive discussion on the theory, algorithms, and applications of quantile regression.

Some recent studies on quantile regression have proposed several methods on analyzing high-dimensional data. Li and Zhu (2008) studied quantile regression with a ℓ_1-norm penalty. They developed an efficient algorithm to compute

* Correspondence to: Department of Mathematics and Statistics, Auburn University, Auburn, AL, 36849, United States
E-mail address: zengpen@auburn.edu (P. Zeng).

https://doi.org/10.1016/j.csda.2019.106819

the solution path of $\boldsymbol{\beta}$ and derived a formula for degrees of freedom, which is used to assess the effective dimension of the fitted model. Wu and Liu (2009) considered quantile regression with a SCAD penalty and proved that the estimates have an oracle property. Belloni and Chernozhukov (2011) considered a scaled ℓ_1-penalized quantile regression (ℓ_1-QR) in high-dimensional sparse models. They studied the theoretical properties of the estimators and showed that ℓ_1-QR is consistent at the near oracle rate under general conditions. Li et al. (2010) studied the penalized quantile regression, including lasso, group lasso and elastic net penalties, from a Bayesian perspective. Galvao and Montes-Rojas (2010) applied the penalized quantile regression for dynamic panel data. They showed that the penalty term reduces the dynamic panel bias and increases the efficiency of the estimators.

The statistical model subject to linear constraints of the coefficients arises commonly in many applications. These constraints are imposed on the parameters to accommodate prior knowledge and assumptions of the problem. Jagannathan and Ma (2003) solved the portfolio variance minimization problem, where portfolio weights are bounded and are also constrained to sum to one. The least squares problem with linear equality constraints are also widely used in signal processing and pattern recognition, see for instance, Li et al. (1995) and Kalouptisidis (1997). Koenker and Ng (2005) proposed an interior point algorithm for solving linear quantile regression subject to linear inequality constraints. It is a modification of the Frisch–Newton algorithm introduced by Portnoy and Koenker (1997). Bollaerts et al. (2006) discussed a non-parametric quantile regression using P-splines which assumes that the estimated quantile function is monotonically increasing. However, there is no existing literature on regularized quantile regression with linear constraints to the best of our knowledge.

To this end, this article will study a regularized quantile regression with linear constraints. Consider estimating α and $\boldsymbol{\beta}$ by solving the following optimization problem,

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \boldsymbol{\beta}^T \boldsymbol{x}_i) + \lambda \|\boldsymbol{D}\boldsymbol{\beta}\|_1, \quad \text{subject to } \boldsymbol{C}\boldsymbol{\beta} \geq \boldsymbol{d} \text{ and } \boldsymbol{E}\boldsymbol{\beta} = \boldsymbol{f}, \tag{1}$$

where $\lambda \geq 0$ is a tuning parameter, $\|\cdot\|_1$ is the ℓ_1-norm of a vector, $\boldsymbol{D} \in \mathbb{R}^{m\times p}$, $\boldsymbol{C} \in \mathbb{R}^{q\times p}$, $\boldsymbol{d} \in \mathbb{R}^{q}$, $\boldsymbol{E} \in \mathbb{R}^{s\times p}$, and $\boldsymbol{f} \in \mathbb{R}^{s}$ are constant matrices or vectors specified by users according to assumptions or subject matter knowledge in application. Problem (1) becomes the ℓ_1-norm quantile regression in Li and Zhu (2008) when $\boldsymbol{D}$ is selected to be the identity matrix and there is no linear constraints. Various choices of $\boldsymbol{D}$ in (1) yield adaptive lasso, fused lasso, or other variants. The linearly constrained generalized lasso with squared-error loss has been studied by He (2011), James et al. (2013) and Hu et al. (2015), etc. But they are not robust compared with quantile regression in the presence of noise and outliers.

Problem (1) proposes a general framework of quantile regression for high-dimensional data with linear constraints. The solution algorithm is a big challenge because the check loss function and penalty are nondifferentiable. What is more, the components of $\boldsymbol{\beta}$ are connected with the constraints and the penalty, which makes the problem more difficult than the existing ℓ_1-norm quantile regression. The main contribution of this paper lies in the following aspects.

(i) In terms of computing, an efficient algorithm is developed to compute $\{(\alpha(\lambda), \boldsymbol{\beta}(\lambda)), 0 \leq \lambda \leq \infty\}$, the whole solution path of problem (1), instead of simply obtaining a solution at a given λ. The Karush–Kuhn–Tucker (KKT) conditions of the problem is derived. It is also shown that the solution path is piecewise constant with respect to λ.

(ii) In terms of theory, a formula for the number of degrees of freedom is derived based on Stein's framework, which can be used as a measure of the model complexity. Two model selection criteria, namely SIC and GACV, are studied and compared for their performance on model selection.

The rest of the paper is organized as follows. Section 2 derives the KKT conditions of the optimization problem and describes the properties of the solution. Section 3 studies the entire solution path of the parameters as a function of λ. In Section 4, a formula of degrees of freedom is obtained using Stein's unbiased risk estimation (Stein, 1981). Simulation studies in different scenarios and two real data examples are conducted to illustrate the application of the proposed method in Sections 5 and 6, respectively. Some conclusion remarks are included in Section 7. All technical proofs of lemmas and theorems are postponed to Appendix.

Notation convention: $\boldsymbol{X} = (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T)^T \in \mathbb{R}^{n\times p}$ is the design matrix and $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ is the response vector. $\boldsymbol{X}_k$ is the kth row of $\boldsymbol{X}$. For an index set $\mathcal{E}$, $\boldsymbol{X}_\mathcal{E}$ is the matrix obtained by selecting the rows of $\boldsymbol{X}$ with the indexes in $\mathcal{E}$. The same rule applies to matrices $\boldsymbol{D}$, $\boldsymbol{C}$ and index sets $\mathcal{A}$, $\mathcal{B}$. $\boldsymbol{y}_\mathcal{E}$ is the vector with elements y_i for $i \in \mathcal{E}$. The same rule applies to vectors $\boldsymbol{v}$, $\boldsymbol{u}$, $\boldsymbol{\xi}$, $\boldsymbol{d}$ and index sets $\mathcal{E}$, $\mathcal{A}$, $\mathcal{B}$. $\mathbf{0}_p$ is a p-dimensional vector of zeros. The same rule applies to $\mathbf{0}_{|\mathcal{E}|}$ and $\mathbf{0}_{|\mathcal{A}|}$, where $|\mathcal{E}|$ and $|\mathcal{A}|$ are the cardinalities of sets $\mathcal{E}$ and $\mathcal{A}$, respectively. $\mathbf{1}_n$ is an n-dimensional vector of ones. The same rule applies to $\mathbf{1}_{|\mathcal{E}|}$.

2. KKT conditions and properties of solution path

This section discusses the properties of the solution $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ to the problem (1). We first formulate the problem (1) as a linear programming problem, and then derive its KKT conditions. The KKT conditions completely characterize $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$, and also shed light on the properties of the whole solution path, that is, $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ as a function of tuning parameter λ.