Computational Statistics and Data Analysis 84 (2015) 27–40

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

An exact polynomial time algorithm for computing the least trimmed squares estimate

Karel Klouda

Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00, Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form 26 October 2014
Accepted 3 November 2014
Available online 8 November 2014

Keywords:
LTS exact algorithm
LTS objective function
Robust estimation

ABSTRACT

An exact algorithm for computing the estimates of regression coefficients given by the least trimmed squares method is presented. The algorithm works under very weak assumptions and has polynomial complexity. Simulations show that in the case of two or three explanatory variables, the presented algorithm is often faster than the exact algorithms based on a branch-and-bound strategy whose complexity is not known. The idea behind the algorithm is based on a theoretical analysis of the respective objective function, which is also given.

1. Introduction

In general, linear regression analysis is concerned with problems of the following type. One random variable Y, called a *response variable*, is supposed to fit the *linear regression model* $Y = \mathbf{x}^T\boldsymbol{\beta}^0 + e$, where $\mathbf{x} \in \mathbb{R}^p$ is a (column) vector of *explanatory variables* (random or otherwise), $\boldsymbol{\beta}^0 \in \mathbb{R}^p$ is a vector of *regression coefficients*, and e is an *error term*. The aim of regression analysis is to estimate $\boldsymbol{\beta}^0$ using the knowledge of n measurements of Y and $\mathbf{x}$. These measurements are denoted as a vector $\mathbf{Y} = (y_1, \ldots, y_n)^T$ and as a *design matrix*

$$\mathbf{X} = \begin{pmatrix} x_1^1 & x_1^2 & \ldots & x_1^p \\ x_2^1 & x_2^2 & \ldots & x_2^p \\ \vdots & \vdots & & \vdots \\ x_n^1 & x_n^2 & \ldots & x_n^p \end{pmatrix}. \tag{1}$$

The vector $\mathbf{x}_i$ stands for the transposition of the ith row of the matrix $\mathbf{X}$.

The best-known estimate of $\boldsymbol{\beta}^0$ is the estimate given by the (ordinary) least squares method (the OLS estimate)

$$\hat{\boldsymbol{\beta}}^{(OLS,n)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{2}$$

which is the projection of $\mathbf{Y}$ into the linear envelope of the columns of $\mathbf{X}$. Unfortunately, the OLS estimate was shown to be very sensitive with respect to data contamination of many types (see Rousseeuw and Leroy, 1987 for details). Therefore, other estimates that are less sensitive or, in other words, more *robust* were introduced. One such estimate is the *least trimmed squares* (LTS) estimate proposed by Rousseeuw (1984).

E-mail address: karel.klouda@fit.cvut.cz.

http://dx.doi.org/10.1016/j.csda.2014.11.001
0167-9473/

The OLS estimate (2) is obtained as a minimum of the OLS objective function (OLS-OF) defined as a sum of squares of *residuals* $r_i(\boldsymbol{\beta}) = y_i - \mathbf{x}_i^T\boldsymbol{\beta}$, i.e., the OLS-OF reads

$$OF^{(OLS,\mathbf{X},\mathbf{Y})}(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2. \tag{3}$$

The LTS method is based on the fact that the contaminating data points typically lay outside the main bulk of data points and hence have larger residuals. To obtain a more robust estimate of the regression coefficients, we ignore (trim) some portion of the data points with the largest residuals. Formally, the LTS estimate is defined as a minimum of the LTS objective function (LTS-OF)

$$OF^{(LTS,n,h)}(\boldsymbol{\beta}) = \sum_{i=1}^{h} r_{(i)}^2(\boldsymbol{\beta}), \tag{4}$$

where h is a parameter that determines how many, namely, $n - h$, data points are to be trimmed and $r_{(i)}^2(\boldsymbol{\beta})$ stands for the ith smallest squared residuum at $\boldsymbol{\beta}$. Because it is not reasonable to ignore more than half of the data points, h usually takes values between $n/2$ and n.

As presented below, there exists a straightforward algorithm that always produces the exact value of the LTS estimate, but it requires $\binom{n}{h}$ computations of OLS estimates for h non-trimmed data points. Because this algorithm is too exhaustive, other, faster algorithms were introduced to provide the estimations.

Most of these algorithms are probabilistic, i.e., it is not certain whether they will return the exact value of the LTS estimate. Two types of probabilistic algorithms exist that may be described, using terminology from Hawkins and Olive (1999), as algorithms for determining $\boldsymbol{\beta}$ satisfying the *weak* and *strong* necessary condition, respectively. Moreover, $\boldsymbol{\beta}$ satisfies the weak necessary condition if and only if it is a point of a local minimum of the LTS-OF (see Theorem 7). Algorithms that determine $\boldsymbol{\beta}$ satisfying the weak necessary condition have been independently proposed several times. The first such algorithm was proposed by Víšek (1996), and its modification can be found in Víšek (2000). Another algorithm of this type was introduced along with the notion of weak necessary condition by Hawkins and Olive (1999), and a version for large data sets was described in Rousseeuw and Van Driessen (1999). For the strong necessary condition, the situation is simple because there is only one representative algorithm: the Feasible Solution Algorithm by Hawkins (1994).

Exact algorithms are presented in Agulló (2001); Hofmann et al. (2010). They are based on the branch-and-bound strategy employed to reduce the number $\binom{n}{h}$ of h-element data subsets for which the OLS estimate must be computed to obtain the exact LTS estimate. The simulation results show that the branch-and-bound strategy is very effective and that the reduction in the average-case complexity is significant. However, the branch-and-bound strategy, in general, does not lower the worst-case complexity, which remains proportional to $\binom{n}{h}$. We use a different strategy to reduce the number of OLS estimate evaluations: we prove that to find the exact LTS estimate, one must compute at most $2^p\binom{n}{p+1}\binom{p}{[p/2]}$ OLS estimates.

Because we are studying an algorithm that solves the problem of minimising the LTS-OF, we can do away with all of the statistical background and formulate the problem as the following optimisation problem:

Problem 1. Find the LTS estimate

$$\hat{\boldsymbol{\beta}}^{(LTS,n,h)} = \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{h} r_{(i)}^2(\boldsymbol{\beta}), \tag{5}$$

where $n > p \geq 1$, $\mathbf{Y} = (y_1, \ldots, y_n)^T$, $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is a matrix from $\mathbb{R}^{n,p}$, and h is an integer such that $p \leq h \leq n$. The data for which the problem is defined is denoted by $\mathcal{D} = \{(y_i, \mathbf{x}_i^T) \mid i \in \{1, \ldots, n\}\}$.

Prior to the introduction of the exact algorithm, we need to study the LTS-OF because the algorithm is based on some of its properties. Having described these properties, we first propose a one-dimensional version of the algorithm, which is convenient for demonstrating the principle of our algorithm. The general case is given afterward.

2. Objective function of the LTS estimate

2.1. Discrete version of LTS-OF

For each $\boldsymbol{\beta} \in \mathbb{R}^p$, only h data points with least squared residuals appear in (4). Every such h-element subset of the data set $\mathcal{D}$ can be unambiguously determined by the $0-1$ vector $\mathbf{w} = (w^1, \ldots, w^n)^T$, where $w^i = 1$ if $(y_i, \mathbf{x}_i^T)$ is an element of this subset and $w^i = 0$ otherwise — in this sense, we speak about a *subset* $\mathbf{w}$. For any element of the set of all such vectors

$$Q^{(n,h)} = \{\mathbf{w} \in \mathbb{R}^n | w^i \in \{0, 1\}, i \in \{1, \ldots, n\}, w^1 + \cdots + w^n = h\}, \tag{6}$$

we define two sets

$$I_{\mathbf{w}} = \{k \in \{1, \ldots, n\} \mid w^k = 1\}, \quad \text{and} \quad O_{\mathbf{w}} = \{k \in \{1, \ldots, n\} \mid w^k = 0\}. \tag{7}$$