Computational Statistics and Data Analysis 71 (2014) 128–137

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Robust mixture regression model fitting by Laplace distribution

Weixing Song [*,1], Weixin Yao, Yanru Xing

Department of Statistics, Kansas State University, Manhattan, KS, 66503, United States

ARTICLE INFO

Article history:
Received 28 November 2012
Received in revised form 28 June 2013
Accepted 28 June 2013
Available online 4 July 2013

Keywords:
Least absolute deviation
EM algorithm
Mixture regression model
Normal mixture
Laplace distribution

ABSTRACT

A robust estimation procedure for mixture linear regression models is proposed by assuming that the error terms follow a Laplace distribution. Using the fact that the Laplace distribution can be written as a scale mixture of a normal and a latent distribution, this procedure is implemented by an EM algorithm which incorporates two types of missing information from the mixture class membership and the latent variable. Finite sample performance of the proposed algorithm is evaluated by simulations. The proposed method is compared with other procedures, and a sensitivity study is also conducted based on a real data set.

Published by Elsevier B.V.

1. Introduction

Least absolute deviation (LAD) regression has been widely used in practice if robust estimation is desired. The research on its computation and theoretical properties is abundant in the literature. A detailed survey on this topic can be found in Dielman (1984, 2005). It is known that the outliers impact more heavily on mixture linear regression models than on the usual linear regression models, since the outliers not only affect the estimation of the regression parameters, but also possibly totally blur the mixture structure. In this paper, LAD will be applied to a class of mixture linear regression models. Simulation studies show that the proposed estimators of the regression coefficients are robust.

To be specific, let X be a p-dimensional vector of explanatory variables and Y be a scalar response variable. The relationship between Y and X is often investigated through a linear regression model. In the mixture linear regression setup, we assume that with probability $\pi_i,\ i = 1, 2, \ldots, g$, (X', Y) comes from one of the following $g \geq 2$ linear regression models

$$Y = X'\beta_i + \sigma_i\varepsilon_i, \quad i = 1, 2, \ldots, g, \tag{1}$$

where $\sum_{i=1}^{g} \pi_i = 1$, the β_i's are unknown p-dimensional vectors of regression coefficients, and the σ_i's are unknown positive scalars. The random errors ε_i's are assumed to be independent of the X_i's. It is commonly assumed that the density functions of ε_i's are members in a location-scale family with means 0 and variances 1. In the following discussion, the design variable X is assumed to be random, but the proposed estimation procedure also works for the fixed design.

If $g = 1$, the LAD estimator of β is the minimizer of the target function $Q(\beta) = \sum_{j=1}^{n} |Y_j - X_j'\beta|$, where $(X_j', Y_j)_{j=1}^{n}$ is a sample from model (1). Many algorithms have been developed in the literature to tackle the minimization problem $\hat{\beta} = \text{argmin}_{\beta} Q(\beta)$, such as linear programming, least angle regression, modified maximum likelihood method by Li and

* Corresponding author.
E-mail address: weixing@ksu.edu (W. Song).

1 The research is supported by the NSF DMS1205276.

0167-9473/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.csda.2013.06.022

Arce (2004), among others. An often adopted but ad-hoc scheme for finding the MLE of β is to obtain the root of the derivative of $Q(\beta)$. Here σ^2 is treated as a nuisance parameter. By doing this, we obtain

$$\frac{\partial Q(\beta)}{\partial \beta} = -\sum_{j=1}^{n} X_j \operatorname{sgn}(Y_j - X_j'\beta) = 0, \tag{2}$$

where $\operatorname{sgn}(\cdot)$ is the sign function which takes -1, 0, 1 if the argument is negative, 0, and positive, respectively. Let $w_j = 1/|Y_j - X_j'\beta|$, and rewrite the Eq. (2) as $\sum_{j=1}^{n} w_j X_j (Y_j - X_j'\beta) = 0$. Thus by supplying an initial value β_0 for β, the updated value β can be found by the weighted least square solution

$$\beta_1 = \left(\sum_{j=1}^{n} w_j X_j X_j' \right)^{-1} \sum_{j=1}^{n} w_j X_j Y_j, \tag{3}$$

where $w_j = 1/|Y_j - X_j'\beta_0|$. By iterating the procedure, one can eventually find an approximate solution to $\operatorname{argmin}_\beta Q(\beta)$.

A very interesting connection between the iterated weighted least square procedure stated above and an EM algorithm in conjunction with the Laplace distribution is found in Phillips (2002). For the sake of completeness, we briefly describe the procedure proposed in Phillips (2002).

Andrews and Mallows (1974) showed that a Laplace distribution can be expressed as a mixture of a normal distribution and another distribution related to the exponential distribution. To be specific, suppose Z and V are two random variables, V has a distribution with density function $v^{-3}\exp(-(2v^2)^{-1}),\ v > 0$, and given $V = v$, the conditional distribution of Z is normal with mean 0 and variance $\sigma^2/(2v^2)$. Then Z marginally has a Laplace distribution with density function $h_\varepsilon(z) = \exp(-\sqrt{2}|z|/\sigma)/(\sqrt{2}\sigma)$. Based on this, Phillips (2002) developed an EM algorithm to search for the minimizer of $Q(\beta)$.

If V could be observed, then the complete log-likelihood function of $\theta = (\beta, \sigma^2)$, based on the sample $\mathbf{P} = (X_j, Y_j, V_j)_{j=1}^n$, is

$$L(\theta; \mathbf{P}) = -\frac{n}{2}\log(\pi\sigma^2) - \frac{1}{\sigma^2}\sum_{j=1}^{n} V_j^2 (Y_j - X_j'\beta)^2 - \sum_{j=1}^{n} \log V_j^2 - \frac{1}{2}\sum_{j=1}^{n}\frac{1}{V_j^2}.$$

Following the two steps in the EM algorithm procedure, and assuming that $\theta^{(k)} = (\beta^{(k)}, \sigma^{2(k)})$ is the value for the kth iteration, then in the $(k+1)$th iteration, we have to first calculate the conditional expectation of the complete log likelihood function $L(\theta; \mathbf{P})$, given the observed data set $(Y_j, X_j)_{j=1}^n$ and $\theta = \theta^{(k)}$, which has the following form

$$\begin{aligned} E[L(\theta; \mathbf{P})|\mathbf{S}] = & -\frac{n}{2}\log(\pi\sigma^2) - \frac{\sum_{j=1}^{n} E[V_j^2 | \theta^{(k)}, (X_j, Y_j)_{j=1}^n](Y_j - X_j'\beta)^2}{\sigma^2} \\ & - \sum_{j=1}^{n} E[\log V_j^2 | \theta^{(k)}, (X_j, Y_j)_{j=1}^n] - \frac{1}{2}\sum_{j=1}^{n} E\left[\frac{1}{V_j^2} \middle| \theta^{(k)}, (X_j, Y_j)_{j=1}^n \right]. \end{aligned}$$

In the second step, the conditional expectation is maximized over θ. Denote $w_j = E[V_j^2|\theta^{(k)}, (X_j, Y_j)_{j=1}^n]$, and notice that the third and fourth terms on the right hand side do not involve the unknown regression parameters. Therefore, to maximize the above conditional expectation is equivalent to maximize the following terms with respect to θ,

$$-\frac{n}{2}\log\sigma^2 - \frac{\sum_{j=1}^{n} w_j (Y_j - X_j'\beta)^2}{\sigma^2}.$$

Interestingly, Phillips (2002) showed $w_j = E[V_j^2|\theta^{(k)}, (X_j, Y_j)_{j=1}^n] = \sigma^{(k)}/(\sqrt{2}|Y_j - X_j'\beta^{(k)}|)$. This implies that the solution $\beta^{(k+1)}$ is the same as the one based on (3) and the iteratively reweighted least squares procedure is an application of the EM algorithm. It is also easy to see that $\sigma^{2(k+1)}$ can be estimated by $2\sum_{j=1}^{n} w_j (Y_j - X_j'\beta^{(k+1)})^2/n$.

The robustness property of the LAD procedure, and the natural connection between LAD estimation and maximum likelihood estimation for the regression coefficients given Laplace distributed random error when $g = 1$, motivate us to consider the possible extension of the algorithm to the mixture model setup. When $g \geq 2$, we assume that for each i, $i = 1, 2, \ldots, g$, ε_i follows a Laplace distribution with location 0 and scale parameter $1/\sqrt{2}$, which results in the variance of ε_i being 1. Then it is easily seen that for a sample $\mathbf{S} = \{(X_j', Y_j),\ j = 1, 2, \ldots, n\}$ from the model (1), the log-likelihood function of $\theta = (\beta_1, \sigma_1^2, \pi_1, \beta_2, \sigma_2^2, \pi_2, \ldots, \beta_g, \sigma_g^2, \pi_g)$ can be written as

$$L(\theta; \mathbf{S}) = \sum_{j=1}^{n} \log\left[\sum_{i=1}^{g} \frac{\pi_i}{\sqrt{2}\sigma_i} \exp\left(-\frac{\sqrt{2}|Y_j - X_j'\beta_i|}{\sigma_i} \right) \right]. \tag{4}$$

Download English Version:

https://daneshyari.com/en/article/6870312

Download Persian Version:

https://daneshyari.com/article/6870312

Daneshyari.com