Computational Statistics and Data Analysis 96 (2016) 1–11

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

The Expectation–Maximization approach for Bayesian quantile regression

Kaifeng Zhao [a,b], Heng Lian [c,*]

[a] Data and Analytics R&D Group, GroupM, Singapore, 048423, Singapore

[b] Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore, 637371, Singapore

[c] School of Mathematics and Statistics, University of New South Wales, Sydney, NSW, 2052, Australia

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form 5 October 2015
Accepted 9 November 2015
Available online 29 November 2015

Keywords:
Bayesian inference
Expectation–Maximization
Model selection
Quantile regression

ABSTRACT

This paper deals with Bayesian linear quantile regression models based on a recently developed Expectation–Maximization Variable Selection (EMVS) method. By using additional latent variables, the proposed approach enjoys enormous computational savings compared to commonly used Markov Chain Monte Carlo (MCMC) algorithm. Using location-scale mixture representation of asymmetric Laplace distribution (ALD), we develop a rapid and efficient Expectation–Maximization (EM) algorithm, which is illustrated with several carefully designed simulation examples. We further apply the proposed method to construct financial index tracking portfolios.

1. Introduction

Over the past decades, variable selection methods have attracted much attention and played an increasingly important role in statistical research and many related fields as well, ranging from bioinformatics, ecology, economics to finance and so on. Specifically, for high dimensional problems, such as gene selection from microarray data (Zhang et al., 2006) or portfolio construction from a large pool of stocks (George and McCulloch, 1997), the selected submodel usually has improved interpretability, increased predictive ability and higher prediction speed.

Various methods have been developed for variable selection problems. See, for example, the lasso (Tibshirani, 1996), Lars (Efron et al., 2004) and boosting (Bühlmann, 2006). From a Bayesian point of view, George and McCulloch (1997) proposed a Stochastic Search Variable Selection (SSVS) approach based on well-known Markov Chain Monte Carlo (MCMC) algorithm. This method has been quite popular in the Bayesian literature, such as Li and Zhang (2010), Scheipl et al. (2012) and Hu et al. (2013), but is also widely known to suffer from heavy computational burden. Yi et al. (2011) considered posterior mode estimation in Gaussian process regression. Ročková and George (2014) proposed an Expectation–Maximization Variable Selection (EMVS) method. In their work, EMVS is shown to be an accurate deterministic approach with enormous computational savings, and has the potential to be a key player for variable selection problems.

Quantile regression models, on the other hand, are known as important alternatives to mean regression models, and provide more robust estimations and complete descriptions of the underlying distribution of the response variable. They have found numerous applications in the fields of economics, biomedicine and others (Cade and Noon, 2003; Yu et al., 2003; Yoshida, in press). In the Bayesian context for quantile regression, Kozumi and Kobayashi (2011) utilized asymmetric Laplace distribution (ALD) for error terms as in Yu and Moyeed (2011), and proposed an efficient Gibbs sampling algorithm by using a

* Corresponding author.
E-mail address: heng.lian@unsw.edu.au (H. Lian).

http://dx.doi.org/10.1016/j.csda.2015.11.005
0167-9473/

location-scale mixture representation of the ALD. Subsequent Bayesian works further extend this method to various models, including single-index models (Hu et al., 2013), binary models (Benoit et al., 2013), Tobit models (Yue and Hong, 2012; Zhao and Lian, 2015), longitudinal data models (Geraci and Bottai, 2006; Luo et al., 2012) and partial linear additive models (Hu et al., 2015).

Various methods have been developed for Bayesian quantile regression with variable selection (Li et al., 2010). In their work, several penalty-based regularization approaches are successfully adopted to the Bayesian context, including LASSO, group LASSO and elastic net penalties. As an alternative, the spike-and-slab prior-based method has also gained increasing attention, which is initially implemented with MCMC and recently developed in EM framework (Ročková and George, 2014). In this work, we will further extend it for variable selection in quantile regression problems. We demonstrate that the location-scale mixture representation, which was used for MCMC previously, also plays a critical role in deriving the EMVS approach for quantile regression, resulting in a fast implementation of Bayesian quantile regression. Without using this important representation, it is not clear a priori that the EM algorithm can be efficiently implemented for estimating quantiles. To be more specific, for any quantile level $\tau \in (0, 1)$, we are interested in estimating τth conditional quantile of the response variable, which is given by

$$Q_{y_i}(\tau) = \alpha + \boldsymbol{x}_i^T \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{1}$$

where $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$ are independent and identically distributed observations, y_i's are response variables, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ are p-dimensional covariates, α is the intercept and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a vector of parameters.

To enable variable selection and a fast implementation of the EM algorithm, we introduce indicator variables as in the spike-and-slab prior in George and McCulloch (1997). As a result, the quantile regression approach proposed in this paper has simultaneous model estimation and variable selection abilities, the latter of which is achieved based on the estimated posterior probabilities of the indicator variables.

The remainder of the paper proceeds as follows. We formulate our Bayesian hierarchical model, including model setup and choices of priors in Section 2. Detailed EM algorithm derivations are presented in Section 3. Section 4 reports results of simulated examples. The proposed approach is further illustrated with a real data example in Section 5. We conclude our work with a brief discussion in Section 6.

2. Hierarchical Bayesian modelling

2.1. Model setup

For any quantile level $\tau \in (0, 1)$, we are interested in estimating the quantile (1). Note α and $\boldsymbol{\beta}$ implicitly depend on τ.

One Bayesian approach commonly used to implement quantile regression is to write $y_i = \alpha + \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ and assume the error ε_i follows an asymmetric Laplace distribution (ALD), of which the probability density function has the following form,

$$\pi(y|\mu, \delta) = \frac{\tau(1-\tau)}{\delta} \exp\left\{-\frac{1}{\delta}\rho_\tau(y-\mu)\right\},$$

where the quantile level τ is the skewness parameter in ALD, $\delta > 0$ is the scale parameter, μ is the location parameter, $\rho_\tau(\cdot)$ is the check loss function defined by $\rho_\tau(y) = y(\tau - I\{y < 0\})$, and $I\{\cdot\}$ is the indicator function.

Following Kozumi and Kobayashi (2011), we employ the location-scale mixture representation of the ALD and rewrite our model as

$$y_i = \alpha + \boldsymbol{x}_i^T \boldsymbol{\beta} + k_1 e_i + \sqrt{k_2 \delta e_i} z_i,$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $k_2 = \frac{2}{\tau(1-\tau)}$, $z_i \sim N(0, 1)$, $e_i \sim \exp(1/\delta)$ and e_i is independent of z_i. Here $\exp(1/\delta)$ denotes the exponential distribution with mean δ.

2.2. Prior specification

To facilitate Bayesian variable selection, we choose the well-known spike-and-slab priors for β_j's. Indicator variables, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^T$, are introduced such that $\beta_j = 0$ if and only if $\gamma_j = 0$, $j = 1, \ldots, p$.

The prior we assigned to $\boldsymbol{\beta}$ is,

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \delta) = N(\mathbf{0}_p, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \quad \text{with } \boldsymbol{\Sigma}_\beta = \delta \cdot \text{diag}(b_1, b_2, \ldots, b_p),$$

where $b_j = (1 - \gamma_j) \cdot v_0 + \gamma_j \cdot v_1$ for $0 \le v_0 \le v_1$. Even though the v_0 in the spike distribution is typically set to be zero in the literature, such as in Brown et al. (2002), Panagiotelis and Smith (2008) and Hu et al. (2015), we follow Ročková and George (2014) and consider small but positive values for v_0 to exclude unimportant nonzero effects. This approach enables a fast EM algorithm. As suggested by Ročková and George (2014), we impose a heavy-tailed prior for v_1,

$$p(v_1) = \frac{v_1^b (1 + v_1)^{-a-b-2}}{B(a+1, b+1)} I\{v_1 > 0\},$$