Journal of Computational and Applied Mathematics 273 (2015) 251–263

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Parallel maximum likelihood estimator for multiple linear regression models

Guangbao Guo [a,b,*], Wenjie You [c], Guoqi Qian [d], Wei Shao [e]

[a] Department of Statistics, Shandong University of Technology, Zibo 255000, China
[b] School of Mathematics, Shandong University, Jinan 250100, China
[c] Department of Automation, Xiamen University, Xiamen 361005, China
[d] Department of Mathematics and Statistics, The University of Melbourne Parkville VIC 3010, Australia
[e] School of Management, Qufu Normal University, Rizhao 276800, China

ARTICLE INFO

Article history:
Received 23 August 2012
Received in revised form 29 September 2013

MSC:
62J05
65F15
65G50

Keywords:
Multiple linear regression models
Parallel computing
Maximum likelihood estimator
Consistency
Outlier

ABSTRACT

Consistency and run-time are important questions in performing multiple linear regression models. In response, we introduce a new parallel maximum likelihood estimator for multiple linear models. We first provide an equivalent condition between the method and the generalized least squares estimator. We also consider the rank of projections and the eigenvalue. We then present consistency when a stable solution exists. In this paper, we describe several consistency theorems and perform experiments on consistency, outlier, and scalability. Finally, we fit the proposed method onto bankruptcy data.

1. Introduction

Multiple linear regression models (MLRMs) are widely used in many statistical problems. Several parallel methods and specified criterion of chosen subsets have been proposed in recent years (see [1]) to improve run-time and computational effect. We provide a brief overview of parallel methods that are useful for solving the MLRMs.

Mitchell and Beauchamp [2] created a parallel method for the subset selection problem using a Bayesian perspective. Havránek and Stratkoš [3] considered parallel methods for the Cholesky factorization in multiple linear models, and showed that the methods' performance is independent of the size of data sets.

Xu et al. [4] suggested a form of stochastic domain decomposition in multiple linear models to improve performance and to resist processor failure. The general concept of domain decomposition is decomposing the data so that the processors have data sets of nearly the same sizes and computation times. The importance of size during parallel communication was similarly considered.

Skvoretz et al. [5] experimented with MLRMs in social science research. The parallel component of their computation was computing for a covariance matrix through a single program multiple data stream. Different numbers of processors were

* Corresponding author at: Department of Statistics, Shandong University of Technology, Zibo 255000, China.
E-mail address: ggb11111111@163.com (G. Guo).

http://dx.doi.org/10.1016/j.cam.2014.06.005
0377-0427/

used in the experiments, and the varying amounts of data were read from a disk. They found that the latter consideration was critical in obtaining good performance.

Bouyouli et al. [6] developed global minimal and global orthogonal residual methods for MLRMs, all of which were block Krylov subspace methods of parallel methods.

MLRMs allow for a highly effective parallel implementation, elegantly illustrating our point and encouraging further development in theory and application. This work originates from the statistical analysis of multiple linear models [7] in statistical tests and from several examples of parallel maximum likelihood estimator (PMLE). Properties of stochastic domain decomposition were studied for the maximum likelihood estimator (MLE) in multiple linear models.

We provide a general MLE in multiple linear models. Suppose that MLRMs have the following form:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I), \tag{1.1}$$

where $X \in \mathbb{R}^{n\times p}$ is a known matrix of fixed rank, $\text{rank}(X) = p,\ p \ll n$, $Y \in \mathbb{R}^{n\times 1}$ is an observable random vector, $\beta \in \mathbb{R}^{p\times 1}$ is a vector of unknown parameters, $I \in \mathbb{R}^{n\times n}$ is a known unit matrix, and σ^2 is a positive unknown parameter.

The MLE is often used to estimate unknown parameters in multiple linear models. The MLE of β under the model (1.1) is defined as

$$\hat{\beta} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta). \tag{1.2}$$

We then have

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{1.3}$$

The PMLE method in (1.1) is as follows: first, (X, Y) is sent to the r processor respectively; second, different elements of (X, Y) are acquired by stochastic domain decomposition in each processor, denoted as (X_i, Y_i); the MLE is then computed, the PMLE is obtained using the estimator in each processor. The PMLE method is a domain decomposition method, and has a short run-time on large data sets. Although there are a number of existing methods for MLRM, the method is more faster, and more robust in some cases.

We organized the rest of this paper as follows. In Section 2, we introduce the PMLE method (see [8]), and provides an equivalence condition of the PMLE and a generalized least squares estimator. In Section 3, we consider first the rank of projections in the PMLE method, followed by the eigenvalue. We study the consistency of the PMLE in Section 4. In Section 5, we illustrate the method through several experiments studies, including those on consistency, outlier and scalability. Experiments with bankruptcy data are also provided. Section 6 discusses future research. The Appendix lists the technical results.

2. PMLE of MLRMs

In this section, we introduce the matrix form of the proposed PMLE in (1.1). We assume that X_i $(i = 1, \ldots, r)$ are the subsamples of the observed sample X, where $X \in \mathbb{R}^{n\times 1}$. Write

$$X_i = R_i X, \qquad E_i = R_i^T R_i, \qquad E_i = \text{diag}\{\alpha_1, \alpha_2, \ldots, \alpha_n\}, \qquad \text{rank}\{E_i\} = n_0 \geq p, \quad i = 1, \ldots, r. \tag{2.1}$$

Here, R_i is the projection operator. $\sum_{i=1}^n \alpha_i = n_0$, $\alpha_i \sim B(n_0, 1/n)$, $E(\alpha_i) = n_0/n$. Note that

$$I \leq \sum_{i=1}^r E_i \leq qI. \tag{2.2}$$

Here q is the number of matrices E_i with a nonzero in the row.

Let $Y_i = R_i Y$ such that $Y_i = X_i\beta_i + \varepsilon_i$, $E(\varepsilon_i) = 0$, and

$$E(X^T E_i X \beta_i | X) = \frac{n_0}{n} X^T X \beta_i; \qquad E(X^T E_i Y | X) = \frac{n_0}{n} X^T Y, \quad i = 1, \ldots, r.$$

We write

$$\hat{\beta}_i = (X^T E_i X)^- X^T E_i Y, \quad i = 1, \ldots, r. \tag{2.3}$$

Assume that

$$\tilde{\beta} = \frac{1}{r}\sum_{i=1}^r \hat{\beta}_i = \frac{1}{r}\sum_{i=1}^r (X^T E_i X)^- X^T E_i Y, \tag{2.4}$$

which is the PMLE in (1.1). Then the parallel estimator is a generalized least squares (GLS) method with domain decomposition. In order to have a good understanding of the method, we give an illustration of the PMLE method in Fig. 1. Now we give a following remark about the method.

We establish the conditions under which the covariance matrix Σ exists, such that the PMLE is equivalent to the GLS estimator.

$$\hat{\beta}_G = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y, \tag{2.5}$$

where $\Sigma \in \mathbb{R}^{n\times n}$ is a nonsingular matrix.