A Fast EM Algorithm for Gaussian Mixtures

Hyekyung Jung\textsuperscript{a}, Byungtae Seo\textsuperscript{1, b}

\textsuperscript{a} Samsungcard CO., LTD.; \textsuperscript{b} Department of Statistics, Sungkyunkwan University

Abstract

The EM algorithm is the most important tool to obtain the maximum likelihood estimator in finite mixture models due to its stability and simplicity. However, its convergence rate is often slow because the conventional EM algorithm is based on a large missing data space. Several techniques have been proposed in the literature to reduce the missing data space. In this paper, we review existing methods and propose a new EM algorithm for Gaussian mixtures, which reduces the missing data space while preserving the stability of the conventional EM algorithm. The performance of the proposed method is evaluated with other existing methods via simulation studies.

Keywords: EM algorithm, ECM algorithm, constrained Newton method.

1. Introduction

For finite mixture models, the maximum likelihood estimator (MLE) does not have a closed form and requires some numerical strategy to find the MLE. Newton-type optimization algorithms can be used as a general purpose; however, they are unstable and hard to program because mixture models involve a large number of parameters and the mixture likelihood has multiple modes. The expectation-maximization (EM) (Dempster et al., 1977) algorithm would be an easy and stable alternative in this case.

The EM algorithm is notoriously slow in many cases. The convergence rate of the EM algorithm depends mainly on the amount of missing information as illustrated in Dempster et al. (1977). The conventional EM algorithm for mixtures is constructed based on a large missing data space that involves all component membership variables. This results in a large missing data space and leads to a slow convergence. If we could characterize a given problem with a smaller missing data space we would expect a faster convergence. This model reduction technique is well studied in Meng and Rubin (1993) and Liu and Rubin (1994), who deem this type of EM algorithms the expectation-conditional-maximization (ECM) algorithm.

Some variants utilize the ECM algorithm in the mixture literature. Celeux et al. (2001) proposed a component-wise EM algorithm for mixtures (CEMM) that updates each set of component parameters and the corresponding mixing weight at each iteration as a tool to reduce missing data space. The CEMM is also considered a Space-Alternating Generalized EM (SAGE) algorithm (Fessler and Hero, 1994) that updates each component parameter based on the reduced missing data space and the mixing proportions with the complete data space. The SAGE algorithm for the Gaussian mixture models is derived in Celeux et al. (1999). Pilla and Lindsay (1996) merged several components to

\footnotesize
\textsuperscript{1} This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0014607).

\textsuperscript{1} Corresponding author: Assistant Professor, Department of Statistics, Sungkyunkwan University, 53 Myeogryun-dong 3-ga, Chongno-gu, Seoul 110-745, Korea. E-mail: seobt@skku.edu
Hyekyung Jung, Byungtae Seo

reduce missing data space in each CM step, but this can only be used for the estimation of component proportions with known component densities. Liu and Sun (1997) used a different strategy to reduce missing data space called the ECME algorithm. They used the conventional EM algorithm for the parameters in the component densities; however, the mixing weights are updated directly from the observed likelihood instead of the complete likelihood. Since the ECME algorithm does not use any missing data space when it estimates mixing proportions, it is expected to be quite fast. However, the estimation of mixing proportions in the observed likelihood is often unstable, so it loses one of the great advantages of using the conventional EM algorithm in mixtures.

In this paper, we propose a stable ECME algorithm using the constrained Newton method for mixing proportions suggested by Wang (2007). The proposed method can also be accelerated by combining CEMM or SAGE algorithms without significant computing effort. This paper is organized as follows: In Section 2 and Section 3, we give a brief review of the existing algorithms including the conventional EM algorithm and the constrained Newton method. In Section 4, we illustrate the proposed algorithm by combining the SAGE with the constrained Newton method. Some simulation studies and concluding remarks are then given in Section 5 and Section 6.

2. Review of Some Existing Methods

In this section, we briefly review the conventional EM algorithm and its variants for finite mixture models. Let us consider a finite $d$-variate normal mixture model,

$$f(x; \theta) = \sum_{j=1}^{m} p_j f(x; \mu_j, \Sigma_j),$$

where $\theta = (\mu_1, \ldots, \mu_m, \Sigma_1, \ldots, \Sigma_m, p_1, \ldots, p_m)$ with constraints $p_j > 0$ and $\sum p_j = 1$. Throughout this paper, we use bold faced lower and upper case letters to represent a vector and a matrix, respectively. Since typical Newton type algorithms are known to be very unstable and difficult to program, the EM algorithm will be used to obtain the MLE of $\theta$. The EM algorithm has many good properties such as simplicity, stability, and monotone convergence. To construct the conventional EM algorithm, one needs to interpret the mixture model as a component missing problem. That is, we assume that each observation $x_i$ comes from one of the component densities while the component membership is missing. In this case, if we define the component membership indicator $z_{ij}$ as

$$z_{ij} = \begin{cases} 1, & \text{if } x_i \text{ comes from } j^{th} \text{ component}, \\ 0, & \text{elsewhere} \end{cases}$$

the joint density of $(x_i, z_{i1}, \ldots, z_{im})$ can then be expressed as

$$\prod_{j=1}^{m} (p_j f(x_i; \mu_j, \Sigma_j))^{z_{ij}}. \tag{2.1}$$

Now, based on the observed $x_i$ and the unobserved $z_{i1}, \ldots, z_{im}, i = 1, \ldots, n$, the conventional EM algorithm to obtain the MLE of $\theta$ is constructed as follows:

E-step: For the current estimate $\theta^{(t)}$

$$Q(\theta | \theta^{(t)}) = \sum_{i=1}^{n} E \left[ \log \left( \prod_{j=1}^{m} (p_j f(x_i; \mu_j, \Sigma_j))^{z_{ij}} \right) \middle| x_i, \theta^{(t)}, \mu_j^{(t)}, \Sigma_j^{(t)} \right]$$

where $Q(\theta | \theta^{(t)})$ is the expected value of the log-likelihood function with respect to the current estimate $\theta^{(t)}$.