Abstract

We survey the EM algorithm and its Monte Carlo-based extensions.

1 The EM Algorithm

The EM algorithm is a method for analyzing incomplete data which was formalized by Dempster et al. (1977). We begin by discussing a probabilistic framework within which the EM algorithm is often applied. We then present the EM algorithm in detail. Finally, we discuss some limitations of this method. Throughout, we illustrate our presentation with a toy problem based on linear regression with unobserved covariates.

1.1 Example: Linear Regression with an Unobserved Covariate

Consider the scenario where a measured quantity is known to depend linearly on another unobserved, but nevertheless well understood, quantity. For example, something, something, census data. We first present a model for such a scenario, then show how to directly analyze the observed data. Throughout the rest of this document, we will return to this example to illustrate how to perform an analysis when increasing portions of the calculations cannot be performed analytically (awk).

Let $X \sim \mathrm{M}(\mu, \tau^2)$, where $\mu \in \mathbb{R}$ and $\tau > 0$. Let $\varepsilon \sim \mathrm{N}(0, \sigma^2)$ for some $\sigma > 0$, and $Y = X\beta + \varepsilon$ where $\beta \in \mathbb{R}$. We observe an iid sample of Ys, but not their corresponding Xs. We do however, treat μ and τ as known. Our goal is to estimate β and σ from this incomplete data.

Appendix A Likelihood for Linear Regression with Unobserved Covariates

In this appendix, we present details for the analysis of our linear regression example with unobserved covariates. See Section 1.1 for formulation of the model and definition of notation.

A.1 Observed Data Likelihood, Score and Information

The complete data distribution for our model can be written as follows.

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu\beta \\ \mu \end{pmatrix}, \begin{bmatrix} \sigma^2 + \tau^2\beta^2 & \tau^2\beta \\ \tau^2\beta & \tau^2 \end{bmatrix} \right) \tag{1}$$

Since our observed data, Y, is a marginal of the complete data, we can read off the distribution of Y from Expression (1). That is, $Y \sim \text{N}(\mu\beta, \sigma^2 + \tau^2\beta^2)$.

Based on a sample of observed data, $y_1, \ldots, y_n$, our log-likelihood is as follows. For conciseness, let $\theta = (\beta, \sigma)$ be the vector of unknown parameters, and $\eta^2 = \sigma^2 + \tau^2\beta^2$ be the marginal variance of Y. In previous work, I used η for $\mathbb{V}Y$. Make sure I adjust any discussion and **CODE(!!!)** accordingly.

$$\ell(\theta; y) = -\frac{n}{2}\log(2\pi) - n\log(\eta) - \sum_{i=1}^{n} \frac{(y_i - \mu\beta)^2}{2\eta^2} \tag{2}$$

$$\equiv -n\log(\eta) - \sum_{i=1}^{n} \frac{(y_i - \mu\beta)^2}{2\eta^2} \tag{3}$$

Where $\equiv$ denotes equality up to additive constants which do not depend on θ.

The score vector is given by

$$S(\theta; y) = \tag{4}$$

As an aside, I did explore the above model with multiple covariates. Unfortunately, marginalizing out X consists of replacing each observed covariate vector with its mean, μ. This results in linearly dependent observations, so the model is overparameterized. I could probably incorporate an intercept term without introducing the overparameterization problem, but I don't think it's worth the effort. I'm not going to be able to sell anyone on the applicability of my model, and adding a third parameter won't really increase the pedagogical value.

Check for "Citation Needed" before publishing.

References

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1977.