The EM in Variance Components

Benny Yakir

December 23, 2010

For the simplicity of the presentation, we consider the case of two components. Generalizing it to a larger nomber of components is straightforward.

Imagine the existence of two random column vectors X_1 and X_2, each with its own length, that are unobserved. Instead, we observe $Y = A_1X_1 + A_2X_2$, for some known matrices A_1 and A_2. The distribution of Y is determined from the distribution of (X_1, X_2). Our goal is to estimate the parameters of this distribution.

Specifically, consider the case where $X_1 \sim N(0, \sigma_1^2 I_1)$ and $X_2 \sim N(0, \sigma_2^2 I_2)$ and are independent of each other. The unknown parameters in this situation are $\phi = (\sigma_1^2, \sigma_2^2)$, which we want to estimate based on the data Y. The estimator is the MLE and the computation method for obtaining this estimator is the EM.

The full data is $x = (x_1, x_2)$. Had we observed this data we would have obtained the likelihood

$$L(\psi; x) = (2\pi\sigma_1^2)^{-r_1/2} e^{-x_1' x_1/(2\sigma_1^2)} \times (2\pi\sigma_2^2)^{-r_2/2} e^{-x_2' x_2/(2\sigma_2^2)}$$

where r_1 and r_2 are the lengths of the unobserved vectors. The Q function of the EM is therefore

$$\begin{aligned} Q(\tilde{\phi}|\phi) = \mathbb{E}[\log L(\tilde{\psi}; x)|y, \psi] &= -\frac{r_1 + r_2}{2} \log(2\pi) \\ - \frac{r_1}{2} \log(\tilde{\sigma}_1^2) - \frac{1}{2\tilde{\sigma}_1^2} \mathbb{E}[x_1' x_1 | y, \psi] &- \frac{r_2}{2} \log(2\pi\tilde{\sigma}_2^2) - \frac{1}{2\tilde{\sigma}_2^2} \mathbb{E}[x_2' x_2 | y, \psi] \end{aligned}$$

In each iteration this function should be maximized with respect to $\tilde{\phi}$, where ϕ is the current value of the parameter. In this example, the maximization is obtained by maximizing for $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ separately. The first maximizer is obtained by taking a derivative with respect to $\tilde{\sigma}_1^2$ and equating to zero in order to get as the solution the value:

$$\hat{\sigma}_1^2 = \mathbb{E}[x_1' x_1 | y, \psi]/r_1 = \frac{1}{r_1} \sum_{i=1}^{r_1} \mathbb{E}(x_{1i}^2)|y, \psi] \ ,$$

where x_{i1} is the ith component of the vector x_1. For the second component we get

$$\hat{\sigma}_2^2 = \frac{1}{r_2} \sum_{i=1}^{r_2} \mathbb{E}(x_{2i}^2)|y, \psi] \ .$$

The last problem to solve is the computation of terms such as $\mathbb{E}[x_{1i}^2|y, \psi]$ and $\mathbb{E}[x_{2i}^2|y, \psi]$. We show the result for x_{1i}. the computation for x_{i2} is similar. Clearly,

$$\mathbb{E}[x_{1i}^2|y, \psi] = \mathbb{V}\mathrm{ar}[x_{1i}|y, \psi] + (\mathbb{E}[x_{1i}|y, \psi])^2 \ .$$

Hence, all we need to do is compute the conditional expectation and variance of x_{1i}, given y. Recall that $y = A_1 x_1 + A_2 x_2$. Consequently, y is the sum of $A_{1i} x_{1i}$, where A_{1i} is the ith column of A_1, and terms that are independent of x_{1i}. It follows that

$$\begin{pmatrix} x_{1i} \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1^2 A_{1i}' \\ \sigma_1^2 A_{1i} & \Sigma \end{pmatrix} \right)$$

where $\Sigma = \sigma_1^2 A_1 A_1' + \sigma_2^2 A_2 A_2'$ is the variance matrix of y. Using the formula for the computation of the conditional normal expectation and variance we get:

$$\mathbb{E}[x_{1i}|y, \psi] = \sigma_1^2 A_{1i}' \Sigma^{-1} y \ , \quad \mathbb{V}\mathrm{ar}[x_{1i}|y, \psi] = \sigma_1^2 - \sigma_1^4 A_{1i}' \Sigma^{-1} A_{1i} \ .$$