EM Algorithm: Examples

1 Mixture model

Suppose Y has p-component mixture density,

$$f(y|\theta) = \sum_{r=1}^{p} \pi_r f_r(y|\theta)$$

where $\pi_r \in [0,1]$, $\sum_{r=1}^{p} \pi_r = 1$, and θ collects both mixing probabilities and parameters in mixture components. Follow the steps to derive the EM algorithm.

1. Let U be the latent variable indicating which subpopulation Y comes from. The joint likelihood of U, Y given θ is

   $$f(y, u|\theta) = \prod_{r=1}^{p} \{\pi_r f_r(y|\theta)\}^{1(u=r)}.$$

   Derive the log-likelihood function for complete data as

   $$\log f(y, u|\theta) = \sum_{r=1}^{p} 1(u = r)\{\log \pi_r + \log f_r(y|\theta)\}.$$

2. The conditional distribution of U given Y, θ is a discrete distribution. Given another θ^* which is different from θ, calculate $w_r(y, \theta^*) := \mathrm{P}(U = r|Y = y, \theta^*)$ according to the joint distribution $f(y, u|\theta)$,

   $$w_r(y, \theta^*) = \frac{\pi_r^* f_r(y|\theta^*)}{\sum_{s=1}^{p} \pi_s^* f_s(y|\theta^*)}, \quad r = 1, \ldots, p.$$

3. (E-step) When the complete data $(y_1, u_1), \ldots, (y_n, u_n)$ are assumed to be i.i.d., show that

   $$\begin{aligned} Q(\theta, \theta^*) &= \mathrm{E}\{\log f(U, Y|\theta)|Y = y, \theta^*\} \\ &= \sum_{j=1}^{n} \sum_{r=1}^{p} w_r(y_j, \theta^*)\{\log \pi_r + \log f_r(y_j|\theta)\} \\ &= \sum_{r=1}^{p} \left\{ \sum_{j=1}^{n} w_r(y_j, \theta^*) \right\} \log \pi_r + \sum_{r=1}^{p} \sum_{j=1}^{n} w_r(y_j, \theta^*) \log f(y_j|\theta) \end{aligned}$$

4. (M-step) If we assume the mixture components are all Gaussian, i.e., $f_r(y|\theta) = f_r(y|\mu_r, \sigma_r^2)$. Maximize $Q(\theta, \theta^*)$ with respect to θ when fixing θ^*. Use $\dagger$ to indicate the maximizer. Show that

$$\pi_r^\dagger = \frac{1}{n}\sum_{j=1}^{n} w_r(y_j, \theta^*),$$

$$\mu_r^\dagger = \frac{\sum_{j=1}^{n} w_r(y_j, \theta^*) y_j}{\sum_{j=1}^{n} w_r(y_j, \theta^*)},$$

$$\sigma_r^{2\dagger} = \frac{\sum_{j=1}^{n} w_r(y_j, \theta^*)(y_j - \mu_r^\dagger)^2}{\sum_{j=1}^{n} w_r(y_j, \theta^*)}.$$

2 Right censored data

Suppose we have n i.i.d. samples from $N(\theta, 1)$. The first m samples are observed while the last $n-m$ samples are censored at constant a. Denote the observed data as

$$\mathbf{x} = (x_1, \ldots, x_m, \underbrace{a, \ldots, a}_{n-m}).$$

The likelihood function based on the observed data is

$$L(\theta|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}(x_i - \theta)^2\right\}\{1 - \Phi(a-\theta)\}^{n-m}.$$

Let ϕ and Φ be the density and distribution functions for the standard normal, respectively.

1. Let $\mathbf{z} = (z_{m+1}, \ldots, z_n)$ be the missing data which are actually censored. Then the complete data log-likelihood is

$$\log L(\theta|\mathbf{x}, \mathbf{z}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{m}(x_i - \theta)^2 - \frac{1}{2}\sum_{i=m+1}^{n}(z_i - \theta)^2.$$

2. Given another θ^* which is different from θ. The distribution of z given $\mathbf{x}, \theta^*$ is $N(\theta^*, 1)$ truncated at constant a from below

$$f(z|\mathbf{x}, \theta^*) = \frac{\exp\{-(z-\theta^*)^2/2\}}{\sqrt{2\pi}\{1 - \Phi(a - \theta^*)\}}, \quad z > a.$$

3. (E-step) Calculate the expected log-likelihood of the complete data under the randomness from $\mathbf{z}$ given $\mathbf{x}, \theta$

$$Q(\theta, \theta^*) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{m}(x_i - \theta)^2 - \frac{1}{2}\sum_{i=m+1}^{n}\int_a^\infty (z_i - \theta)^2 f(z_i|\mathbf{x}, \theta^*) dz_i$$

4. (M-step) Differentiating with respect to θ yields

$$m(\bar{x} - \theta) + (n - m)\{E(z|\mathbf{x}, \theta^*) - \theta\} = 0.$$

For truncated normal, verify that

$$E\{Z|\mathbf{x}, \theta^*\} = \theta^* + \frac{\phi(a - \theta^*)}{1 - \Phi(a - \theta^*)}.$$

Finally, we have

$$\theta^\dagger = \frac{m}{n}\bar{x} + \frac{n - m}{n}\left\{\theta^* + \frac{\phi(a - \theta^*)}{1 - \Phi(a - \theta^*)}\right\}.$$