STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 12
October 3, 2019

Review: EM Algorithm

- **Algorithm 30 (EM Algorithm)**

  Choose an initial value for θ_0 and set $k = 0$.

  1. **Expectation Step:** Compute

  $$Q(\theta \mid \theta_k, x_o) = E[\log f(x_o, x_m \mid \theta) \mid \theta_k, x_o].$$

  2. **Maximization Step:** Set

  $$\theta_{k+1} = \arg\max_{\theta} Q(\theta \mid \theta_k, x_o).$$

  3. Check convergence. If not converged, set $k = k + 1$ and go to Step 1.

Examples of EM Algorithm

- There are some well known examples of using EM algorithms for finding maximum likelihood estimates of parameters.

- In this section, we discuss two of them, namely, estimation of multinomial parameter(s) and estimation of Gaussian mixture parameters.

Multinomial Case

- Consider a multinomial distribution for $x = (x_1, x_2, x_3, x_4)$ parameterized by θ according to:

$$x \sim M(n;\ 0.5 + 0.25\theta,\ 0.25(1-\theta),\ 0.25(1-\theta),\ 0.25\theta)$$

The likelihood function of θ given the vector x is:

$$f(x \mid \theta) = \binom{n}{x_1\ x_2\ x_3\ x_4}(0.5+0.25\theta)^{x_1}(0.25(1-\theta))^{x_2+x_3}(0.25\theta)^{x_4}$$

where

$$\binom{n}{x_1\ x_2\ x_3\ x_4} = \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}$$

Hence, the log-likelihood function is

$$\log(f(x|\theta)) = x_1 \log(2+\theta) + (x_2+x_3)\log(1-\theta) + x_4 \log(\theta) + \text{const.}$$

Multinomial Case

Define two new random variables y_1 and y_2 such that

$$x_1 = y_1 + y_2 ,$$

where

$(y_1, y_2) \sim \text{Binomial}(x_1; 0.5/(0.5 + 0.25\theta), 0.25\theta/(0.5 + 0.25\theta))$

Set the complete data to be $y = (y_1, y_2, x_2, x_3, x_4)$. The vector y has multinomial distribution according to:

$$y \sim M(n; 0.5, 0.25\theta, 0.25(1-\theta), 0.25(1-\theta), 0.25\theta) .$$

The complete data log-likelihood is:

$$(y_2 + x_4)\log(\theta) + (x_2 + x_3)\log(1-\theta) + \text{const.}$$

E-step

(a) **E-Step:** Its expectation with respect to the density function $f(y_1, y_2|\theta_k, x_1, x_2, x_3, x_4)$ is given by:

$$Q(\theta \mid \theta_k, x) = E(\log f(y \mid \theta) \mid \theta_k, x)$$

$$= \int_y [\log f(y \mid \theta)] P(y \mid \theta_k, x) dy$$

$$= \sum_{y_2=0}^{x_1} ((y_2 + x_4)\log\theta + (x_2 + x_3)\log(1-\theta)) P(y_2 \mid \theta_k, x)$$

$$= (E(y_2 \mid \theta_k, x) + x_4)\log\theta + (x_2 + x_3)\log(1-\theta)$$

We need to compute:

$$E(y_2 \mid \theta_k, x)$$

E-step

At the k-th iterate, $(y_1, y_2) \sim \text{Binomial}(x_1; 2/(2+\theta_k), \theta_k/(2+\theta_k))$

$$E(y_2 \mid \theta_k, x) = \sum_{y_2=0}^{x_1} y_2 \frac{x_1!}{y_2!(x_1-y_2)!}\left(\frac{2}{2+\theta_k}\right)^{x_1-y_2}\left(\frac{\theta_k}{2+\theta_k}\right)^{y_2}$$

$$= \sum_{y_2=1}^{x_1} \frac{x_1!}{(y_2-1)!(x_1-y_2)!}\left(\frac{2}{2+\theta_k}\right)^{x_1-y_2}\left(\frac{\theta_k}{2+\theta_k}\right)^{y_2}$$

$$= \left(\frac{\theta_k}{2+\theta_k}\right)x_1 \sum_{y_2-1=0}^{x_1-1} \frac{(x_1-1)!}{(y_2-1)!((x_1-1)-(y_2-1))!}\cdot$$

$$\left(\frac{2}{2+\theta_k}\right)^{(x_1-1)-(y_2-1)}\left(\frac{\theta_k}{2+\theta_k}\right)^{y_2-1} = \left(\frac{\theta_k}{2+\theta_k}\right)x_1$$

Therefore,

$$Q(\theta \mid \theta_k, x) = \left(\frac{\theta_k}{2+\theta_k}x_1 + x_4\right)\log(\theta) + (x_2 + x_3)\log(1-\theta).$$

M-step

(b) **M-Step:** To perform the maximization step, we maximize $Q(\theta|\theta_k, x)$ by taking its derivative and setting it equal to zero.

Solving for θ_{k+1}, we obtain:

$$\theta_{k+1} = \frac{\frac{\theta_k}{2+\theta_k}x_1 + x_4}{\frac{\theta_k}{2+\theta_k}x_1 + x_2 + x_3 + x_4}.$$

This results in the EM algorithm for finding MLE of θ.

Gaussian Mixture Case

- Remember that the observed data in this case is $x_o = Y$, the observations of mixture variable, and the missing data is $x_m = l$, the set of labels associated with elements of Y.

- The log-likelihood function is given by:

$$\log f(Y \mid \theta) = \sum_{i=1}^{n} \log[\alpha_1 f_1(Y_i \mid \mu_1, \sigma_1^2) + \alpha_2 f_2(Y_i \mid \mu_2, \sigma_2^2)]$$

The complete data log-likelihood function for this case is given as:

$$\log f(Y, l \mid \theta) = \sum_{i=1}^{n} \log f(Y_i, l_i \mid \theta)$$

$$= \sum_{i;l_i=1} \log[\alpha_1 f_1(Y_i \mid \mu_1, \sigma_1^2)] + \sum_{i;l_i=2} \log[\alpha_2 f_2(Y_i \mid \mu_2, \sigma_2^2)]$$

E-step

(a) **E-Step:** This step involves computing the expectation of the log-likelihood function with respect to the probability $P(l|Y, \theta_k)$, where l is an n-vector of labels.

The resulting expectation is given by:

$$
\begin{aligned}
Q(\theta \mid \theta_k, Y) &= E[\log f(l, Y \mid \theta) \mid Y, \theta_k] \\
&= \sum_{i=1}^{n} E[\log f(l_i, Y_i \mid \theta) \mid Y, \theta_k] \\
&= \sum_{i=1}^{n} \left(\sum_{l_i=1}^{2} \log f(l_i, Y_i \mid \theta) P(l_i \mid Y_i, \theta_k) \right) \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{2} \log[\alpha_j f_j(Y_i \mid \mu_j, \sigma_j^2)] P(j \mid Y_i, \theta_k) \right)
\end{aligned}
$$

M-step

- That is,

$$f(\{\mu_j,\sigma_j,\alpha_j\}) = Q(\theta \mid \theta_k, Y)$$

$$= \sum_{i=1}^{n}\left(\sum_{j=1}^{2}\left[\log\alpha_j - \log\sigma_j - \frac{(Y_i-\mu_j)^2}{2\sigma_j^2}\right]P(j \mid Y_i,\theta_k)\right) + const.$$

- Update all the parameters:

$$0 = \frac{\partial f}{\partial \mu_j} = \sum_{i=1}^{n}\frac{2(Y_i-\mu_j)}{2\sigma_j^2}P(j \mid Y_i,\theta_k)$$

Therefore,

$$\mu_j = \sum_{i=1}^{n} Y_i P(j \mid Y_i,\theta_k) \Big/ \sum_{i=1}^{n} P(j \mid Y_i,\theta_k)$$

M-step

- Similarly,

$$0 = \frac{\partial f}{\partial \sigma_j} = \sum_{i=1}^{n} [-\frac{1}{\sigma_j} - \frac{(-2)(Y_i - \mu_j)^2}{2\sigma_j^3}] P(j \mid Y_i, \theta_k)$$

Therefore,

$$\sigma_j^2 = \sum_{i=1}^{n} (Y_i - \mu_j)^2 P(j \mid Y_i, \theta_k) \Big/ \sum_{i=1}^{n} P(j \mid Y_i, \theta_k)$$

- To update α_j, we use a Lagrange multiplier. Let

$$g(\{\alpha_j\}) = f(\{\mu_j, \sigma_j, \alpha_j\}) + \lambda(\alpha_1 + \alpha_2 - 1)$$

Then,

$$0 = \frac{\partial g}{\partial \alpha_j} = \sum_{i=1}^{n} \frac{1}{\alpha_j} P(j \mid Y_i, \theta_k) + \lambda$$

M-step

Therefore,

$$1 = \alpha_1 + \alpha_2 = -\frac{1}{\lambda}[\sum_{i=1}^{n} P(j=1 \mid Y_i, \theta_k) + \sum_{i=1}^{n} P(j=2 \mid Y_i, \theta_k)] = -\frac{n}{\lambda}.$$

That is,

$$\alpha_j = \frac{1}{n}\sum_{i=1}^{n} P(j \mid Y_i, \theta_k)$$

- Finally,

$$P(j \mid Y_i, \theta_k) = P(Y_i, l_i = j \mid \theta_k) \Big/ P(Y_i \mid \theta_k)$$

$$= P(Y_i \mid l_i = j, \theta_k) P(l_i = j \mid \theta_k) \Big/ \sum_{j=1}^{2} P(Y_i \mid l_i = j, \theta_k) P(l_i = j \mid \theta_k)$$

$$= \alpha_j f_j(Y_i \mid \mu_j, \sigma_j^2) \Big/ \sum_{j=1}^{2} \alpha_j f_j(Y_i \mid \mu_j, \sigma_j^2).$$

M-step

- (b) **M-Step:** Maximizing the expected log-likelihood function $Q(\theta \mid \theta_k, Y)$, we obtain the following updates:

$$\alpha_{j,k+1} = \frac{1}{n}\sum_{i=1}^{n} P(j \mid Y_i, \theta_k)$$

$$\mu_{j,k+1} = \sum_{i=1}^{n} Y_i P(j \mid Y_i, \theta_k) \Big/ \sum_{i=1}^{n} P(j \mid Y_i, \theta_k)$$

$$\sigma^2_{j,k+1} = \sum_{i=1}^{n} (Y_i - \mu_{j,k+1})^2 P(j \mid Y_i, \theta_k) \Big/ \sum_{i=1}^{n} P(j \mid Y_i, \theta_k)$$

where

$$P(j \mid Y_i, \theta_k) = \alpha_{j,k} f_j(Y_i \mid \mu_{j,k}, \sigma^2_{j,k}) \Big/ \sum_{j=1}^{2} \alpha_{j,k} f_j(Y_i \mid \mu_{j,k}, \sigma^2_{j,k}).$$