BIOSTAT 880 HW3 Solution, Fall 2024

Solution compiled on October 5, 2024

Q 8.8

$$f(y_i \mid k, \theta_i) = \frac{y_i^{k-1}}{\theta_i^k} \exp\left(-\frac{y_i}{\theta_i}\right) \cdot \frac{1}{\Gamma(k)}$$

$$\theta_i = \frac{1}{k} g\left(\sum_j \beta_j x_{ij}\right)$$

The joint likelihood for the complete data is:

$$p(y \mid \eta) = \prod_{i=1}^{n} f(y_i \mid k, \theta_i)$$

where the parameter vector $\eta = (\beta_1, ..., \beta_J, k)$ is $J + 1$ dimensional.
Taking the log, we get:

$$\log p(y \mid \eta) = \sum_{i=1}^{n} \left((k-1)\log y_i - \frac{y_i}{\theta_i} - k \log \theta_i - \log \Gamma(k)\right)$$

For the complete data, the log-likelihood becomes:

$$\log p(y \mid \eta) = \sum_{i=1}^{n} \left(-\frac{y_i}{\theta_i} + (k-1)\log y_i - k \log \theta_i\right) - n \log \Gamma(k)$$

Define $A(\eta)$ as:

$$A(\eta) = -k \sum_{i=1}^{n} \log \theta_i - n \log \Gamma(k)$$

When k is known, we have $\theta_i = \frac{1}{k} g\left(\sum_j \beta_j x_{ij}\right)$, so:

$$l(\eta|\mathbf{y}) = \log p(\mathbf{y} \mid \eta) = \sum_{i=1}^{n} \left(-\frac{y_i}{\theta_i} + (k-1)\log y_i\right) + A(\eta)$$

E-Step: Taking the expectation with respect to the missing data. Assume the current parameter estimate is $\eta^{(t)}$.

$$E_{y_{(1)}}[l(\eta|\mathbf{y}) \mid y_{(0)}, \eta^{(t)}]$$

$$= \sum_{i=1}^{r} \frac{1}{k} g\left(\sum_j \beta_j x_{ij}\right) y_{(0)i} + \sum_{i=1}^{n-r} \frac{1}{k} g\left(\sum_j \beta_j x_{ij}\right) E\left[y_{(1)i} \mid y_{(0)}, \eta^{(t)}\right]$$

$$+(k-1)\sum_{i=1}^{r} \log y_{(0)i} + \sum_{i=1}^{n-r} E[\log y_{(1)i} \mid y_{(0)}, \eta^{(t)}] + A(\eta)$$

where:

$$E[y_{(1)i}] = E[y_i \mid y_i > c]$$

$$E[\log y_{(1)i}] = E[\log y_i \mid y_i > c]$$

by the assumption that y_i is missing iff $y_i > c$.

Then calculate the expectation and plug back to be the estimated full-data likelihood, maximize the likelihood, and obtain $\eta^{(t+1)}$. It can then be worked out in detail.

Q 8.16

Let X be Bernoulli with $\Pr(X = 1) = \pi$ and $\Pr(X = 0) = 1 - \pi$. Also, let $Y \mid X = j$ be normally distributed with mean μ_j and variance σ^2. The full data consists of both X and Y in which $n - r$ values of X are missing.

Full Data Likelihood

The likelihood of the complete data (X, Y), considering that we observe all n values, can be written as:

$$L(\theta \mid X, Y) = \prod_{i=1}^{n} p(X_i, Y_i \mid \theta)$$

This can be separated into observed and missing parts as follows:

$$L(\theta \mid X, Y) = \prod_{i=1}^{r} p(X_i, Y_i \mid \theta) \times \prod_{i=r+1}^{n} p(X_i, Y_i \mid \theta)$$

For the observed part (i.e., $i = 1, \ldots, r$), the likelihood is:

$$P(X_i = 1, Y_i \mid \theta) = \pi \mathcal{N}(Y_i \mid \mu_1, \sigma^2)$$

$$P(X_i = 0, Y_i \mid \theta) = (1 - \pi) \mathcal{N}(Y_i \mid \mu_0, \sigma^2)$$

For the missing part (i.e., $i = r+1, \ldots, n$), we need to integrate over the missing X_i's, which we do implicitly in the E-step.

Taking the log of the full data likelihood, we get:

$$\log L(\theta \mid X, Y) = \sum_{i=1}^{r} \log p(X_i, Y_i \mid \theta) + \sum_{i=r+1}^{n} \log p(X_i, Y_i \mid \theta)$$

For the observed part:

$$\log p(X_i, Y_i \mid \theta) = X_i \log \pi + (1 - X_i) \log(1 - \pi) + X_i \log \mathcal{N}(Y_i \mid \mu_1, \sigma^2) + (1 - X_i) \log \mathcal{N}(Y_i \mid \mu_0, \sigma^2)$$

For the missing part, we replace the missing X_i's by their posterior expectation in the E-step.

E-Step: Expectation of the Full-Data Log-Likelihood

In the E-step, we take the expectation of the log-likelihood with respect to the posterior distribution of the missing X_i's. The posterior distribution for each missing X_i is:

$$\gamma_i^{(t)} = P(X_i = 1 \mid Y_i, \theta^{(t)}) = \frac{\pi^{(t)} \mathcal{N}(Y_i \mid \mu_1^{(t)}, \sigma^{2(t)})}{\pi^{(t)} \mathcal{N}(Y_i \mid \mu_1^{(t)}, \sigma^{2(t)}) + (1 - \pi^{(t)}) \mathcal{N}(Y_i \mid \mu_0^{(t)}, \sigma^{2(t)})}$$

Where $\mathcal{N}$ denotes the p.d.f of the normal Y. Since X_i is a Bernoulli random variable with $\Pr(X_i = 1) = \pi$, we have the following for the expectation of X_i:

$$E[X_i] = \pi$$

However, in the context of the EM algorithm, for missing values of X_i, we compute the posterior expectation $E[X_i \mid \theta^{(t)}, Y_i]$, given the observed data Y_i and the current parameter estimates $\theta^{(t)}$.

The posterior probability that $X_i = 1$ given the observed data Y_i and current estimates $\theta^{(t)}$ is denoted by:

$$\gamma_i^{(t)} = P(X_i = 1 \mid Y_i, \theta^{(t)})$$

Thus, the posterior expectation of X_i is:

$$E[X_i \mid Y_i, \theta^{(t)}] = \gamma_i^{(t)}$$

where $\gamma_i^{(t)}$ is given above.

Thus, the full expected log-likelihood becomes:

$$E[\log L(\theta \mid X, Y)] = \sum_{i=1}^{r} \left[X_i \log \pi + (1 - X_i) \log(1 - \pi) + X_i \log \mathcal{N}(Y_i \mid \mu_1, \sigma^2) + (1 - X_i) \log \mathcal{N}(Y_i \mid \mu_0, \sigma^2) \right]$$

$$+ \sum_{i=r+1}^{n} \left[\gamma_i^{(t)} \log \pi + (1 - \gamma_i^{(t)}) \log(1 - \pi) + \gamma_i^{(t)} \log \mathcal{N}(Y_i \mid \mu_1, \sigma^2) + (1 - \gamma_i^{(t)}) \log \mathcal{N}(Y_i \mid \mu_0, \sigma^2) \right]$$

M-Step

In the M-step of the EM algorithm, we maximize the expected log-likelihood with respect to the parameters π, μ_1, μ_0, and σ^2. Below are the equations for the derivatives of the expected log-likelihood with respect to each parameter.

1. Derivative with respect to π, use x_i to denote the observed data:

$$\frac{\partial}{\partial \pi} \left(\sum_{i=1}^{r} [x_i \log \pi + (1 - x_i) \log(1 - \pi)] + \sum_{i=r+1}^{n} \left[\gamma_i^{(t)} \log \pi + (1 - \gamma_i^{(t)}) \log(1 - \pi) \right] \right) = 0$$

2. Derivative with respect to μ_1:

$$\frac{\partial}{\partial \mu_1}\left(\sum_{i=1}^{r} x_i \frac{-(Y_i-\mu_1)^2}{2\sigma^2} + \sum_{i=r+1}^{n} \gamma_i^{(t)} \frac{-(Y_i-\mu_1)^2}{2\sigma^2}\right) = 0$$

3. Derivative with respect to μ_0:

$$\frac{\partial}{\partial \mu_0}\left(\sum_{i=1}^{r} (1-x_i) \frac{-(Y_i-\mu_0)^2}{2\sigma^2} + \sum_{i=r+1}^{n} (1-\gamma_i^{(t)}) \frac{-(Y_i-\mu_0)^2}{2\sigma^2}\right) = 0$$

4. Derivative with respect to σ^2:

$$\frac{\partial}{\partial \sigma^2}\left(\sum_{i=1}^{r}\left[x_i \frac{(Y_i-\mu_1)^2}{\sigma^2} + (1-x_i)\frac{(Y_i-\mu_0)^2}{\sigma^2}\right] + \sum_{i=r+1}^{n}\left[\gamma_i^{(t)} \frac{(Y_i-\mu_1)^2}{\sigma^2} + (1-\gamma_i^{(t)})\frac{(Y_i-\mu_0)^2}{\sigma^2}\right]\right) = 0$$

Solve the above equations, we can obtain $\theta^{(t+1)} = (\pi^{(t+1)}, \mu_0^{(t+1)}, \mu_1^{(t+1)}, (\sigma^2)^{(t+1)})$ and then use it to calculate the next E-step i.e. $\gamma_i^{(t+1)} = P(X_i = 1 \mid Y_i, \theta^{(t+1)})$. The details were omitted here but you should work out the closed-form expression for each parameter in the midterm.