CSE584 HOMEWORK 5

CHEN SUN
CXS1031@PSU.EDU

1. Question 1

(1)Observed variables: $x_1, ..., x_n$
(2)Missing random variables:

$$\text{mean cluster of point } x_j\text{, denote it as a vector } A_j = \begin{cases} [1,0] & \text{if mean cluster is 1} \\ [0,1] & \text{if mean cluster is 2} \end{cases}$$

$$\text{variance cluster of point } x_j\text{, denote it as a vector } B_j = \begin{cases} [1,0] & \text{if variance cluster is 1} \\ [0,1] & \text{if variance cluster is 2} \end{cases}$$

(3)Unknown parameters: $\pi_1, \pi_2, \gamma_1, \gamma_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.

Complete data likelihood function:

$$p = \prod_{j=1}^{n} \prod_{i=1}^{2} \prod_{k=1}^{2} \left[\frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_j-\mu_i)^2}{2\sigma_k^2}} \pi_i \gamma_k \right]^{A_j[i] \times B_j[k]}$$

Full data loglikelihood function:

$$p = \sum_{j=1}^{n} \sum_{i=1}^{2} \sum_{k=1}^{2} A_j[i] B_j[k] \left[-\log \sigma_k - \frac{(x_j-\mu_i)^2}{2\sigma_k^2} + \log \pi_i + \log \gamma_k \right]$$

Λ function:

$$\begin{aligned} \Lambda = & \sum_{A_1 \ldots A_n B_1 \ldots B_n} q_1(A_1) \ldots q_n(A_n) s_1(B_1) \ldots s_n(B_n) \sum_{j=1}^{n} \sum_{i=1}^{2} \sum_{k=1}^{2} A_j[i] B_j[k] \left[-\log \sigma_k - \frac{(x_j-\mu_i)^2}{2\sigma_k^2} + \log \pi_i + \log \gamma_k \right] \\ & - \sum_{A_1 \ldots A_n B_1 \ldots B_n} q_1(A_1) \ldots q_n(A_n) s_1(B_1) \ldots s_n(B_n) \left[\log q_1(A_1) \ldots q_n(A_n) s_1(B_1) \ldots s_n(B_n) \right] \end{aligned}$$

Simplifying steps:

$$
\begin{aligned}
&\sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\left[\log q_1(A_1)s_1(B_1)\right] \\
&= \sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\left[\log(q_1(A_1)s_1(B_1))\right] \\
&= \sum_{A_1 B_1} q_1(A_1)s_1(B_1)\log(q_1(A_1)s_1(B_1))\left[\sum_{A_2\dots A_n B_2\dots B_n} q_2(A_2)\dots q_n(A_n)s_2(B_2)\dots s_n(B_n)\right] \\
&= \sum_{A_1 B_1} q_1(A_1)s_1(B_1)\log(q_1(A_1)s_1(B_1))
\end{aligned}
$$

So:

$$
\begin{aligned}
&\sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\left[\log q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\right] \\
&= \sum_{A_1 B_1} q_1(A_1)s_1(B_1) + \dots + \sum_{A_n B_n} q_n(A_n)s_n(B_n)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\sum_{j=1}^{n}\sum_{i=1}^{2}\sum_{k=1}^{2} A_j[i]B_j[k]\left[-\log\sigma_k - \frac{(x_j-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&= \sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\sum_{i=1}^{2}\sum_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&\quad + \sum_{A_1\dots A_n B_1\dots B_n} q_1(A_1)\dots q_n(A_n)s_1(B_1)\dots s_n(B_n)\sum_{i=1}^{2}\sum_{k=1}^{2} A_2[i]B_2[k]\left[-\log\sigma_k - \frac{(x_2-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&\quad + \dots \\
&= \sum_{A_1 B_1} q_1(A_1)s_1(B_1)\sum_{i=1}^{2}\sum_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&\quad + \sum_{A_2 B_2} q_2(A_2)s_2(B_2)\sum_{i=1}^{2}\sum_{k=1}^{2} A_2[i]B_2[k]\left[-\log\sigma_k - \frac{(x_2-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&\quad + \dots
\end{aligned}
$$

So the original Λ function can be simplified as:

$$
\begin{aligned}
&\sum_{A_1 B_1} q_1(A_1)s_1(B_1)\sum_{i=1}^{2}\sum_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] - \sum_{A_1 B_1} q_1(A_1)s_1(B_1)\log(q_1(A_1)s_1(B_1)) \\
&+ \dots \\
&+ \sum_{A_n B_n} q_n(A_n)s_n(B_n)\sum_{i=1}^{2}\sum_{k=1}^{2} A_n[i]B_n[k]\left[-\log\sigma_k - \frac{(x_n-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] - \sum_{A_n B_n} q_n(A_n)s_n(B_n)\log(q_n(A_n)s_n(B_n))
\end{aligned}
$$

There are $2n$ formulations in Λ function, considering the first two formulation:
$\sum\limits_{A_1 B_1} q_1(A_1)s_1(B_1)\sum\limits_{i=1}^{2}\sum\limits_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] - \sum_{A_1 B_1} q_1(A_1)s_1(B_1)\log(q_1(A_1)s_1(B_1))$

Denote $h(A_1 B_1) = \sum\limits_{i=1}^{2}\sum\limits_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right]$

By KL-divergence trick, $q_1(A_1)s_1(B_1) = \dfrac{e^{h(A_1B_1)}}{\sum\limits_{A_1B_1} e^{h(A_1B_1)}}$

So the probability under $q \cdot s$ that first point has mean from cluster i and deviation from cluster k is:

$$T_1[i,k] = \frac{\frac{\pi_i\gamma_k}{\sigma_k}e^{-\frac{(x_1-\mu_i)^2}{2\sigma_k^2}}}{\sum\limits_{\iota=1}^{2}\sum\limits_{\kappa=1}^{2}\frac{\pi_\iota\gamma_\kappa}{\sigma_\kappa}e^{-\frac{(x_1-\mu_\iota)^2}{2\sigma_\kappa^2}}}$$

For the first formulation of Λ function, plugging in:

$$\begin{aligned}
&\sum_{A_1B_1} q_1(A_1)s_1(B_1)\sum_{i=1}^{2}\sum_{k=1}^{2} A_1[i]B_1[k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] \\
&= \sum_{i=1}^{2}\sum_{k=1}^{2}\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right]\sum_{A_1B_1} q_1(A_1)s_1(B_1)A_1[i]B_1[k] \\
&= \sum_{i=1}^{2}\sum_{k=1}^{2} T_1[i,k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right]
\end{aligned}$$

q **update.** So after q update, the Λ function is:

$$\begin{aligned}
\Lambda = &\sum_{i=1}^{2}\sum_{k=1}^{2} T_1[i,k]\left[-\log\sigma_k - \frac{(x_1-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] + \text{something} \\
&+ \sum_{i=1}^{2}\sum_{k=1}^{2} T_2[i,k]\left[-\log\sigma_k - \frac{(x_2-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] + \text{something} \\
&+ \dots
\end{aligned}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{2}\sum_{k=1}^{2} T_j[i,k]\left[-\log\sigma_k - \frac{(x_j-\mu_i)^2}{2\sigma_k^2} + \log\pi_i + \log\gamma_k\right] + \text{something, where } T_j[i,k] = \frac{\frac{\pi_i\gamma_k}{\sigma_k}e^{-\frac{(x_j-\mu_i)^2}{2\sigma_k^2}}}{\sum\limits_{\iota=1}^{2}\sum\limits_{\kappa=1}^{2}\frac{\pi_\iota\gamma_\kappa}{\sigma_\kappa}e^{-\frac{(x_j-\mu_\iota)^2}{2\sigma_\kappa^2}}}$$

μ_i **update.** Considering μ_i:

$$\begin{aligned}
\mu_i &= argmax \sum_{j=1}^{n}\sum_{k=1}^{2} T_j[i,k]\left[-\frac{(x_j-\mu_i)^2}{2\sigma_k^2}\right] \\
\frac{\partial}{\partial\mu_i} &= \sum_{j=1}^{n}\sum_{k=1}^{2}\frac{T_j[i,k]}{\sigma_k^2}(x_j-\mu_i) \\
\mu_i &= \frac{\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{2}\frac{T_j[i,k]}{\sigma_k^2}x_j}{\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{2}\frac{T_j[i,k]}{\sigma_k^2}}
\end{aligned}$$

π **update.**

$$\pi = argmax \sum_{j=1}^{n}\sum_{i=1}^{2}\sum_{k=1}^{2} T_j[i,k]\log\pi_i, \text{ such that } \sum_{i=1}^{2}\pi_i = 1$$

Introducing lagrange multiplier λ:

$$\pi = argmax \sum_{j=1}^{n} \sum_{i=1}^{2} \sum_{k=1}^{2} T_j[i,k] \log \pi_i + \lambda(\sum_{i=1}^{2} \pi_i - 1)$$

$$\frac{\partial}{\partial \pi_1} = \sum_{j=1}^{n} \sum_{k=1}^{2} \frac{T_j[1,k]}{\pi_1} + \lambda = 0$$

$$\text{so: } \pi_1 \lambda + \sum_{j=1}^{n} \sum_{k=1}^{2} T_j[1,k] = 0$$

$$\frac{\partial}{\partial \pi_2} = \sum_{j=1}^{n} \sum_{k=1}^{2} \frac{T_j[2,k]}{\pi_2} + \lambda = 0$$

$$\text{so: } \pi_2 \lambda + \sum_{j=1}^{n} \sum_{k=1}^{2} T_j[2,k] = 0$$

Add them up:

$$\lambda + \sum_{j=1}^{n} \sum_{i=1}^{2} \sum_{k=1}^{2} T_j[i,k] = 0$$

Since we have $\sum_{i=1}^{2} \sum_{k=1}^{2} T_j[i,k] = 1$, so $\lambda + n = 0$.
So $\lambda = -n$.
So $\pi_1 = \frac{\sum_{j=1}^{n} \sum_{k=1}^{2} T_j[1,k]}{n}$; $\pi_2 = \frac{\sum_{j=1}^{n} \sum_{k=1}^{2} T_j[2,k]}{n}$.

σ_k^2 **update.**

$$\begin{aligned}
\text{let } t_k &= \sigma_k^2 \\
t_k &= argmax \sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,k] \left[-\log \sqrt{t_k} - \frac{(x_j - \mu_i)^2}{2t_k} \right] \\
\frac{\partial}{\partial t_k} &= \sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,k] \left[-\frac{1}{2t_k} + \frac{(x_j - \mu_i)^2}{2t_k^2} \right] = 0 \\
\sigma_k^2 = t_k &= \frac{\sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,k](x_j - \mu_i)^2}{\sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,k]}
\end{aligned}$$

γ **update.** Similar as π update, we have: So $\gamma_1 = \frac{\sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,1]}{n}$; $\gamma_2 = \frac{\sum_{j=1}^{n} \sum_{i=1}^{2} T_j[i,2]}{n}$.