Stat 246 Homework III - Solution

CA: Darongsae Kwon

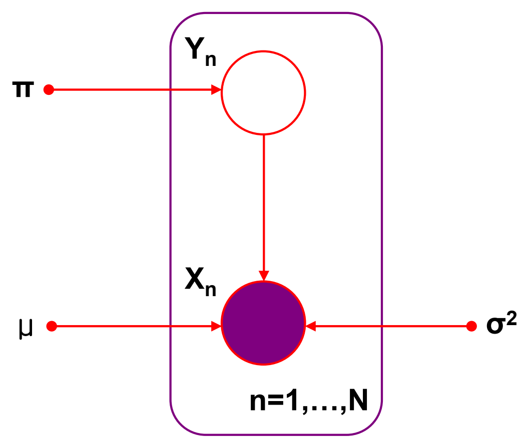
1 [50 points]

We have a mixture of M univariate Gaussians.

$$f(x;\theta) = \sum_{m=1}^{M} \pi_m \phi(x; \mu_m, \sigma_m^2),$$

where $\phi(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$.

(a) [5 points]



(b) [10 points]

The joint distribution of (X, Y) is

$$\begin{aligned} f(x,y) &= \prod_{m=1}^{M} \left[\pi_m \phi(x;\mu_m,\sigma_m^2)\right]^{I[y=m]} \\ &= \prod_{m=1}^{M} \left[\pi_m \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left\{-\frac{1}{2\sigma_m^2}(x-\mu_m)^2\right\}\right]^{I[y=m]}, \end{aligned}$$

so the log-likelihood is

$$l(\theta) = \sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\log \pi_m - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\log \sigma_m^2 - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\frac{(X_n - \mu_m)^2}{\sigma_m^2}.$$

Since we have the constraint $\sum_{m=1}^{M} \pi_m = 1$, we maximize $l(\theta)$ with a Lagrange multiplier. Define

$$\Lambda(\theta) = \sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\log \pi_m - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\log \sigma_m^2 - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m]\frac{(X_n - \mu_m)^2}{\sigma_m^2} - \lambda\sum_{m=1}^{M}\pi_m.$$

Then, the MLE of θ can be obtained by maximizing $\Lambda(\theta)$ as follows.

$$\begin{aligned}
\frac{\partial \Lambda}{\partial \pi_m} &= \sum_{n=1}^{N} I[Y_n = m]\frac{1}{\pi_m} - \lambda = 0 \\
\Rightarrow \qquad \hat{\pi}_m &= \frac{1}{\lambda}\sum_{n=1}^{N} I[Y_n = m] \overset{(*)}{=} \frac{1}{N}\sum_{n=1}^{N} I[Y_n = m],
\end{aligned}$$

where the equality (*) follows from $\sum_{m=1}^{M} \hat{\pi}_m = 1$,

$$\begin{aligned}
\frac{\partial \Lambda}{\partial \mu_m} &= \sum_{n=1}^{N} I[Y_n = m]\frac{X_n - \mu_m}{\sigma_m^2} = 0 \\
\Rightarrow \qquad \hat{\mu}_m &= \frac{\sum_{n=1}^{N} I[Y_n = m]X_n}{\sum_{n=1}^{N} I[Y_n = m]} = \frac{\sum_{n=1}^{N} I[Y_n = m]X_n}{N\hat{\pi}_m}, \\
\frac{\partial \Lambda}{\partial \sigma_m^2} &= -\frac{1}{2\sigma_m^2}\sum_{n=1}^{N} I[Y_n = m] + \frac{1}{2(\sigma_m^2)^2}\sum_{n=1}^{N} I[Y_n = m](X_n - \mu_m)^2 = 0 \\
\Rightarrow \qquad \hat{\sigma}_m^2 &= \frac{\sum_{n=1}^{N} I[Y_n = m](X_n - \hat{\mu}_m)^2}{\sum_{n=1}^{N} I[Y_n = m]} = \frac{\sum_{n=1}^{N} I[Y_n = m](X_n - \hat{\mu}_m)^2}{N\hat{\pi}_m}.
\end{aligned}$$

If we don't have any observations such that $Y_n = m$, then $\hat{\pi}_m = 0$, so we cannot estimate μ_m and σ_m^2.

Note: Here, you can also use the 1-of-M representation for Y as shown in Problem 2 below. In that case, $I[Y_n = m]$ above is replaced with $Y_{m,n}$.

(c) [5 points]

The log-likelihood function for $M = 2$ with $\mu_1, \mu_2, \pi_1, \pi_2$ as given is

$$\begin{aligned}
&l(\sigma_1^2, \sigma_2^2) \\
= \; &\sum_{n=1}^{N}\log\left(\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{1}{2\sigma_1^2}(X_n - X_1)^2\right\} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left\{-\frac{1}{2\sigma_2^2}(X_n - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\right\}\right) \\
= \; &\log\left(\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma_1^2}} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left\{-\frac{1}{2\sigma_2^2}(X_1 - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\right\}\right) \\
&+ \sum_{n=2}^{N}\log\left(\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{1}{2\sigma_1^2}(X_n - X_1)^2\right\} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left\{-\frac{1}{2\sigma_2^2}(X_n - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\right\}\right).
\end{aligned}$$

Note that $\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma_1^2}} \to \infty$ and $\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{1}{2\sigma_1^2}(X_n - X_1)^2\right\} \to 0$ for $n = 2, \cdots, N$ as $\sigma_1^2 \to 0$. Thus, for any fixed σ_2, the first log term in the log-likelihood function goes to infinity while each of the remaining log terms goes to some fixed constant. This makes the entire log-likelihood function go to infinity as $\sigma_1^2 \to 0$ and it is not possible to find the maximum likelihood estimate of σ_1^2 in this case.

(d) [5 points]

If we assume $\sigma_1 = \sigma_2 = \sigma$,

$$\begin{aligned} l(\sigma^2) &= \log\Bigg[\Bigg(\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma^2}} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_1 - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\}\Bigg) \\ &\quad \times \prod_{n=2}^{N}\Bigg(\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_n - X_1)^2\} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_n - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\}\Bigg)\Bigg]. \end{aligned}$$

Again, as $\sigma^2 \to 0$, we have

$$\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma^2}} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_1 - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\} \to \infty, \tag{1}$$

$$\frac{1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_n - X_1)^2\} + \frac{N-1}{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(X_n - \frac{1}{N-1}\sum_{k=2}^{N}X_k)^2\} \to 0, \tag{2}$$

for $n = 2, \cdots, N$. Note that the product of the quantities in (1) and (2) goes to 0 as $\sigma^2 \to 0$, since the convergence in (2) is faster than in (1). Thus, the likelihood goes to 0 as $\sigma^2 \to 0$. The likelihood goes to 0 as $\sigma^2 \to \infty$ as well. Thus, we can find the maximum likelihood estimate of σ^2 in this case.

(e) [10 points]

Under the setting in (d), we can write out the log-likelihood function with a Lagrange multiplier as

$$\Lambda(\theta) = \sum_{n=1}^{N}\sum_{m=1}^{2} I[Y_n = m]\log\pi_m - \frac{N}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{2} I[Y_n = m](X_n - \mu_m)^2 - \lambda\sum_{m=1}^{M}\pi_m.$$

Then, $\hat{\pi}_m$ and $\hat{\mu}_m$ are the same as in (b) and $\hat{\sigma}^2$ is obtained as follows.

$$\begin{aligned} \frac{\partial\Lambda}{\partial\sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}\sum_{m=1}^{2} I[Y_n = m](X_n - \mu_m)^2 = 0 \\ \Rightarrow \qquad \hat{\sigma}^2 &= \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} I[Y_n = m](X_n - \hat{\mu}_m)^2. \end{aligned}$$

(f) [10 points]

First, compute $q_n^{(t)}(y)$.

$$\begin{aligned} q_n^{(t)}(y) &= P(y|X_n; \theta^{(t)}) \\ &= \frac{\pi_y^{(t)}\phi(X_n; \mu_y^{(t)}, \sigma^{2(t)})}{\sum_{z=1}^{M}\pi_z^{(t)}\phi(X_n; \mu_z^{(t)}, \sigma^{2(t)})}. \end{aligned}$$

From the MLE of $\theta = (\mu_1, \mu_2, \sigma^2, \pi_1, \pi_2)$ given in (e), we can derive $\theta^{(t+1)}$ as follows.

$$
\begin{aligned}
\pi_m^{(t+1)} &= \frac{1}{N}\sum_{n=1}^{N}\sum_{y=1}^{M} q_n^{(t)}(y) I[y=m] \\
&= \frac{1}{N}\sum_{n=1}^{N} q_n^{(t)}(m), \\
\mu_m^{(t+1)} &= \frac{1}{N\pi_m^{(t+1)}}\sum_{n=1}^{N}\sum_{y=1}^{M} q_n^{(t)}(y) I[y=m] X_n \\
&= \frac{1}{N\pi_m^{(t+1)}}\sum_{n=1}^{N} q_n^{(t)}(m) X_n, \\
\sigma^{2(t+1)} &= \frac{1}{N}\sum_{n=1}^{N}\sum_{y=1}^{M} q_n^{(t)}(y) \sum_{m=1}^{M} I[y=m](X_n - \mu_m^{(t+1)})^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} q_n^{(t)}(m)(X_n - \mu_m^{(t+1)})^2.
\end{aligned}
$$

Then, the EM algorithm is

Initialize Set $t = 0$.
1) Assign a label Y_n from $\{1, \cdots, M\}$ randomly to each observation X_n, $n = 1, \cdots, N$ and assume that we have a set of full observations, (X_n, Y_n) , $n = 1, \cdots, N$.
2) Compute the MLEs for the full data as given in (e).
3) If $\hat{\pi}_m = 0$, then set $\hat{\pi}_m = \epsilon$, a small positive number. If $\hat{\pi}_m = 1$, then set $\hat{\pi}_m = 1 - \epsilon$.
4) Set $\theta^{(0)} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, \hat{\pi}_1, \hat{\pi}_2)$.

E Step Given $\theta^{(t)}$,
1) compute

$$
q_n^{(t)}(m) = \frac{\pi_m^{(t)} \phi(X_n; \mu_m^{(t)}, \sigma^{2(t)})}{\sum_{k=1}^{M} \pi_k^{(t)} \phi(X_n; \mu_k^{(t)}, \sigma^{2(t)})}.
$$

2) If $q_n^{(t)}(m) = 0$, then set $q_n^{(t)}(m) = \epsilon$, a small positive number. If $q_n^{(t)}(m) = 1$ then set $q_n^{(t)}(m) = 1 - \epsilon$.

M Step Given $q_n^{(t)}(m)$,
1) compute

$$
\begin{aligned}
\pi_m^{(t+1)} &= \frac{1}{N}\sum_{n=1}^{N} q_n^{(t)}(m), \\
\mu_m^{(t+1)} &= \frac{1}{N\pi_m^{(t+1)}}\sum_{n=1}^{N} q_n^{(t)}(m) X_n, \\
\sigma^{2(t+1)} &= \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} q_n^{(t)}(m)(X_n - \mu_m^{(t+1)})^2.
\end{aligned}
$$

Computation of $L^{(t+1)}$ Given $\theta^{(t+1)}$,
compute the marginal log-likelihood as

$$\begin{aligned} L^{(t+1)} &= \sum_{n=1}^{N} \log f(X_n; \theta^{(t+1)}) \\ &= \sum_{n=1}^{N} \log \left\{ \sum_{m=1}^{M} \pi_m^{(t+1)} \phi(X_n; \mu_m^{(t+1)}, \sigma^{2(t+1)}) \right\}. \end{aligned}$$

Stopping Rule Given TOL,
stop the iteration if $L^{(t+1)} - L^{(t)} < TOL$.

(g) [5 points]

Assume that $\mu_1 = \mu_2$ and let $\mu_1 = \mu_2 = \frac{1}{N}\sum_{n=1}^{N} X_n$ with $\pi_1 = \frac{1}{N}$ and $\pi_2 = \frac{N-1}{N}$. Then, we can avoid the degeneracy in (c). Or, you can put priors for σ_1 and σ_2.

2 [20 points]

The marginal distribution of X is

$$p(x; \theta) = \sum_{m=1}^{M} \pi_m f_m(x; \theta_m),$$

where $f_m(x; \theta_m) = \prod_{\alpha=1}^{d} p_{\alpha,m}^{x_\alpha}(1 - p_{\alpha,m})^{1-x_\alpha}$.

Introduce a discrete random variable Y having M possible states as in Problem 1. Using the 1-of-M representation, the joint distribution of (X, Y) is given by

$$\begin{aligned} p(x, y; \theta) &= \prod_{m=1}^{M} [\pi_m f_m(x; \theta_m)]^{y_m} \\ &= \prod_{m=1}^{M} \left[\pi_m \prod_{\alpha=1}^{d} p_{\alpha,m}^{x_\alpha}(1 - p_{\alpha,m})^{1-x_\alpha} \right]^{y_m}, \end{aligned}$$

where y_m is the m-th coordinate of y and can only take 0 or 1.

The log-likelihood for the full observations $(X_1, Y_1), \ldots, (X_N, Y_N)$ is

$$l(\theta) = \sum_{n=1}^{N} \sum_{m=1}^{M} \left[Y_{m,n} \log \pi_m + \sum_{\alpha=1}^{d} \{ Y_{m,n} X_{\alpha,n} \log p_{\alpha,m} + Y_{m,n}(1 - X_{\alpha,n}) \log(1 - p_{\alpha,m}) \} \right],$$

where $X_{\alpha,n}$ is the α-th coordinate of X_n and $Y_{m,n}$ is the m-th coordinate of Y_n.

Since we have the constraint $\sum_{i=1}^{M} \pi_m = 1$, we maximize $l(\theta)$ by the method of Lagrange multipliers. Define

$$\Lambda(\theta) = \sum_{n=1}^{N} \sum_{m=1}^{M} \left[Y_{m,n} \log \pi_m + \sum_{\alpha=1}^{d} \{ Y_{m,n} X_{\alpha,n} \log p_{\alpha,m} + Y_{m,n}(1 - X_{\alpha,n}) \log(1 - p_{\alpha,m}) \} \right] - \lambda \sum_{m=1}^{M} \pi_m.$$

Then, the MLE of θ is obtained by maximizing Λ.

$$\frac{\partial \Lambda}{\partial \pi_m} = \sum_{n=1}^{N} Y_{m,n}/\pi_m - \lambda = 0$$

$$\Rightarrow \qquad \hat{\pi}_m = \frac{1}{\lambda}\sum_{n=1}^{N} Y_{m,n} = \frac{1}{N}\sum_{n=1}^{N} Y_{m,n},$$

where the last equality follows from $\sum_{m=1}^{M} \hat{\pi}_m = 1$,

$$\frac{\partial \Lambda}{\partial p_{\alpha,m}} = \sum_{n=1}^{N}\left[Y_{m,n}X_{\alpha,n}\frac{1}{p_{\alpha,m}} - Y_{m,n}(1-X_{\alpha,n})\frac{1}{1-p_{\alpha,m}}\right] = 0$$

$$\Rightarrow \qquad \hat{p}_{\alpha,m} = \frac{\sum_{n=1}^{N} Y_{m,n}X_{\alpha,n}}{\sum_{n=1}^{N} Y_{m,n}} = \frac{1}{N\hat{\pi}_m}\sum_{n=1}^{N} Y_{m,n}X_{\alpha,n}.$$

The $q_n^{(t)}(m)$ in the EM algorithm is derived as in Problem 1.

$$\begin{aligned} q_n^{(t)}(m) &= P(Y_{m,n}=1|X_n;\theta^{(t)}) \\ &= \frac{\pi_m^{(t)} f_m(X_n;\theta_m{}^{(t)})}{\sum_{k=1}^{M}\pi_k^{(t)} f_k(X_n;\theta_k^{(t)})} \\ &= \frac{\pi_m^{(t)}\prod_{\alpha=1}^{d} p_{\alpha,m}^{(t)X_{\alpha,n}}(1-p_{\alpha,m}^{(t)})^{1-X_{\alpha,n}}}{\sum_{k=1}^{M}\pi_k^{(t)}\prod_{\alpha=1}^{d} p_{\alpha,k}^{(t)X_{\alpha,n}}(1-p_{\alpha,k}^{(t)})^{1-X_{\alpha,n}}}. \end{aligned}$$

As in Problem 1, we can compute the estimated parameters for the EM algorithm from the MLE equations for the full data.

$$\begin{aligned} \pi_m^{(t+1)} &= \frac{1}{N}\sum_{n=1}^{N} q_n^{(t)}(m), \\ p_{\alpha,m}^{(t+1)} &= \frac{1}{N\pi_m^{(t+1)}}\sum_{n=1}^{N} q_n^{(t)}(m)X_{\alpha,n}, \end{aligned}$$

In summary, the EM algorithm for the given mixture Model is

Initialize Set $t = 0$.

1) Assign a label Y_n from M possible states randomly to each observation X_n, $n = 1, \cdots, N$ and assume that we have a set of full observations (X_n, Y_n) , $n = 1, \cdots, N$.

2) Compute the MLEs for the full data as follows.

$$\begin{aligned} \hat{\pi}_m &= \frac{1}{N}\sum_{n=1}^{N} Y_{m,n}, \\ \hat{p}_{\alpha,m} &= \frac{1}{N\hat{\pi}_m}\sum_{n=1}^{N} Y_{m,n}X_{\alpha,n}. \end{aligned}$$

3) If $\hat{p}_{\alpha,m} = 0$ or $\hat{\pi}_m = 0$, then set $p_{\alpha,m} = \epsilon$ or $\hat{\pi}_m = \epsilon$, where ϵ is a small positive number. If $\hat{p}_{\alpha,m} = 1$ or $\hat{\pi}_m = 1$, then set $p_{\alpha,m} = 1 - \epsilon$ or $\hat{\pi}_m = 1 - \epsilon$.

4) Set $\theta^{(0)} = (\hat{p}_{\alpha,m}, \hat{\pi}_m)$, $\alpha = 1, \cdots, d$ and $m = 1, \cdots, M$.

E Step Given $\theta^{(t)}$,
1) compute $q_n^{(t)}(m)$. Instead of multiplying small numbers, work with the exponent of sums of logs. Let

$$\begin{aligned} r_n^{(t)}(m) &= \log \left\{ \pi_m^{(t)} \prod_{\alpha=1}^{d} p_{\alpha,m}^{(t) X_{\alpha,n}} (1 - p_{\alpha,m}^{(t)})^{1-X_{\alpha,n}} \right\} \\ &= \log \pi_m^{(t)} + \sum_{\alpha=1}^{d} \left\{ X_{\alpha,n} \log p_{\alpha,m}^{(t)} + (1 - X_{\alpha,n}) \log(1 - p_{\alpha,m}^{(t)}) \right\}. \end{aligned}$$

Then,

$$\begin{aligned} q_n^{(t)}(m) &= \frac{\pi_m^{(t)} \prod_{\alpha=1}^{d} p_{\alpha,m}^{(t) X_{\alpha,n}} (1 - p_{\alpha,m}^{(t)})^{1-X_{\alpha,n}}}{\sum_{k=1}^{M} \pi_k^{(t)} \prod_{\alpha=1}^{d} p_{\alpha,k}^{(t) X_{\alpha,n}} (1 - p_{\alpha,k}^{(t)})^{1-X_{\alpha,n}}} \\ &= \exp \left\{ r_n^{(t)}(m) - \log \sum_{k=1}^{M} \exp r_n^{(t)}(k) \right\}. \end{aligned}$$

2) If $q_n^{(t)}(m) = 0$, then set $q_n^{(t)}(m) = \epsilon$. If $q_n^{(t)}(m) = 1$ then set $q_n^{(t)}(m) = 1 - \epsilon$.

M Step Given $q_n^{(t)}$,
1) compute

$$\begin{aligned} \pi_m^{(t+1)} &= \frac{1}{N} \sum_{n=1}^{N} q_n^{(t)}(m), \\ p_{\alpha,m}^{(t+1)} &= \frac{1}{N \pi_m^{(t+1)}} \sum_{n=1}^{N} q_n^{(t)}(m) X_{\alpha,n}. \end{aligned}$$

2) If $\pi_m^{(t+1)} = 0$ or $p_{\alpha,m}^{(t+1)} = 0$, then set $\pi_m^{(t+1)} = \epsilon$ or $p_{\alpha,m}^{(t+1)} = \epsilon$ with a small positive number ϵ. If $\pi_m^{(t+1)} = 1$ or $p_{\alpha,m}^{(t+1)} = 1$, then $\pi_m^{(t+1)} = 1 - \epsilon$ or $p_{\alpha,m}^{(t+1)} = 1 - \epsilon$.

Computation of $L^{(t+1)}$ Given $\theta^{(t+1)}$,
compute the marginal log-likelihood as

$$\begin{aligned} L^{(t+1)} &= \sum_{n=1}^{N} \log f(X_n; \theta^{(t+1)}) \\ &= \sum_{n=1}^{N} \log \left\{ \sum_{m=1}^{M} \pi_m^{(t+1)} \prod_{\alpha=1}^{d} p_{\alpha,m}^{(t+1) X_{\alpha,n}} (1 - p_{\alpha,m}^{(t+1)})^{1-X_{\alpha,n}} \right\} \\ &= \sum_{n=1}^{N} \log \sum_{m=1}^{M} \exp r_n^{(t+1)}(m). \end{aligned}$$

Stopping Rule Given TOL,
stop the iteration if $L^{(t+1)} - L^{(t)} < TOL$.

3 [30 points]

(a)

The following is an R code to implement the EM algorithm.

```
### initialize parameters
init.para=function(X,M,tol){
    d=dim(X)[2]
    N=dim(X)[1]
    #1) assign Yn randomly from M states
    Yn=sample(1:M, N, replace=T)
    #2) compute the MLEs
    Pi.m=double(M)
    P.am=double(d*M)
    for(m in 1:M){
        Pi.m[m]= sum(Yn==m)/N
        P.am[(1+(m-1)*d):(m*d)]= colSums(X[Yn==m,])/(N*Pi.m[m])
    }
    #3)check if there is zero or one in the estimates and replace them with epsilon or 1-epsilon
    Pi.m[Pi.m < tol]=tol
    Pi.m[Pi.m > (1-tol)]=1-tol
    P.am[P.am < tol]=tol
    P.am[P.am > (1-tol)]=1-tol
    return(c(P.am, Pi.m))
}
### compute r=log of unnormalized q
Qn.t=function(Xn, theta){
    d=length(Xn)
    M=length(theta)/(d+1)
    Pa=matrix(theta[1:(d*M)],M,d,byrow=T)
    Pim=theta[(d*M+1):((d+1)*M)]
    res=double(M)
    for(m in 1:M){
        res[m]=sum(Xn*log(Pa[m,]) + (1-Xn)*log(1-Pa[m,]) ) + log(Pim[m])
    }
    return(res)
}
### compute q, theta, the marginal log-likelihood in each EM step
EMstep=function(X, theta,tol){
    d=dim(X)[2]
    N=dim(X)[1]
    M=length(theta)/(d+1)
    res=double(M*(d+1))
    #store r=log of unnormalized q
    Qnmat=matrix(0,N,M)
    Qnmat2=matrix(0,N,M)
    for(n in 1:N){
```

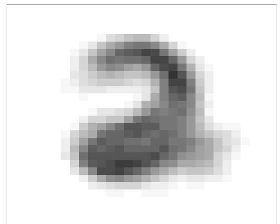
```
        Qnmat[n,]=Qn.t(X[n,],theta)
    }
    #compute q from r=log of unnormalized q
    minQ=apply(Qnmat,c(1),FUN=min)
    temp=Qnmat-minQ
    temp[temp > 700]=700
    Qnmat2= exp(temp)/rowSums(exp(temp))
    Qnmat2[Qnmat2 < tol]=tol
    Qnmat2[Qnmat2 > (1-tol)]=1-tol
    #estimate Pi.m
    res[(M*d+1):(M*d+M)]=colSums(Qnmat2)/N
    #estimate P.am
    for(m in 1:M){
        Q=Qnmat2[,m]%*%matrix(1,1,d)
        res[(1+(m-1)*d):(m*d)]=(colSums(Q*X)/N)/res[(M*d+m)]
    }
    res[res < tol]=tol
    res[res > (1-tol)]=1-tol
    #estimate Lt
    Lik=sum(minQ+log(rowSums(exp(temp))))
    return(list(par=res,logLik=Lik))
}
### run the EM algorithm
EM=function(X, M, tol){
    dist=1
    step=0
    lik.old=-1e5
    lik.new=-1e5
    param=init.para(X,M,tol) #store initialized estimates
    while(dist > 1e-5){
        tt=EMstep(X,param,tol)
        param=tt$par
        lik.new=tt$logLik

        dist=lik.new-lik.old #compute increase in log-likelihood
        lik.old=lik.new
        step=step+1
    }
    return(list(theta=param, logLik=lik.new, iterations=step))
}

M=3
est=EM(twos,M,1e-15)
theta=est$theta
```

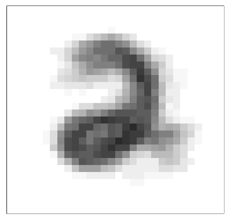
(b) See Next Page.

M=1

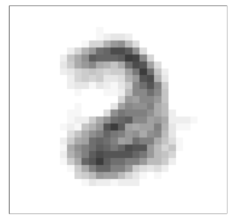


m=1

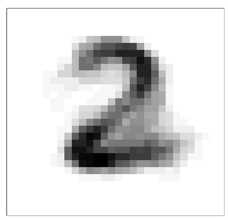
M=3



m=1

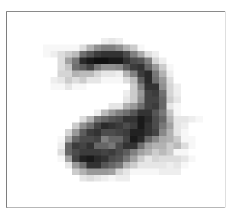


m=2

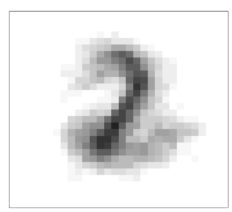


m=3

M=5

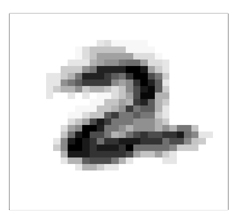


m= 1

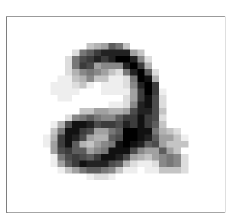


m= 2

m= 3



m= 4



m= 5