University of California, Davis

STA 243

Homework 3

Student:	***Student ID:***
Bohao Zou	**917796070**
Bingdao Chen	**917781027**

May 25, 2020

Pledge:

Please sign below (print full name) after checking (✓) the following. If you can not honestly check each of these responses, please email me at kbala@ucdavis.edu to explain your situation.

- We pledge that we are honest students with academic integrity and we have not cheated on this homework.
- These answers are our own work.
- We did not give any other students assistance on this homework.
- We understand that to submit work that is not our own and pretend that it is our is a violation of the UC Davis code of conduct and will be reported to Student Judicial Affairs.
- We understand that suspected misconduct on this homework will be reported to the Office of Student Support and Judicial Affairs and, if established, will result in disciplinary sanctions up through Dismissal from the University and a grade penalty up to a grade of "F" for the course.

Team Member 1 :
Bohao Zou

Team Member 2 :
Bingdao Chen

Q1

Please read the code of R.

Q2

(a)

The marginal distribution of the Z_i is

$$P(Z_i = j) = \pi_j,\ j = 1, 2, \cdots, k$$

(b)

$$\begin{aligned}
P(Z_i = j|x_i) &= \frac{P(x_i|Z_i = j)P(Z_i = j)}{P(x_i)} \\
&= \frac{P(x_i|Z_i = j)P(Z_i = j)}{\sum_{j=1}^{k} P(x_i|Z_i = j)P(Z_i = j)} \\
&= \frac{\frac{\pi_j}{(2\pi)^{\frac{d}{2}}|\sum_j|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_i-\mu_j)^\top \sum_j^{-1}(x_i-\mu_j)\right)}{\sum_{j=1}^{k}\frac{\pi_j}{(2\pi)^{\frac{d}{2}}|\sum_j|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_i-\mu_j)^\top \sum_j^{-1}(x_i-\mu_j)\right)}
\end{aligned}$$

(c)

$$\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{n}\log\left[\sum_{j=1}^{k} F_{ij}\frac{P_\theta(x_i, Z_i = j)}{F_{ij}}\right] \\
&= \sum_{i=1}^{n}\log\left[\mathbb{E}_{q_i}\left(\frac{P_\theta(x_i, Z_i = j)}{F_{ij}}\right)\right] \\
&\geq \sum_{i=1}^{n}\mathbb{E}_{q_i}\left(\log\left[\frac{P_\theta(x_i, Z_i = j)}{F_{ij}}\right]\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{k} F_{ij}\log\left[\frac{P_\theta(x_i, Z_i = j)}{F_{ij}}\right]
\end{aligned}$$

By applying nsen's inequality $\log\mathbb{E}X \geq \mathbb{E}\log X$, we complete the proof.

(d)

If $\ell(\theta^{'}) = Q(F, \theta^{'})$, then the equality of Jensen's inequality holds, $\log\mathbb{E}(X) = \mathbb{E}\log X$. For Jensen's inequality, equality holds if and only if function $f(x)$ is strong convex and $P(X = \mathbb{E}X) = 1$. In this case, $f(x) = -\log x$.
First, $f(x) = -\log x$ is convex because $f^{''}(x) = \frac{1}{x^2} > 0$. Then, we prove that $X = \mathbb{E}X$. In this case, $X = \frac{P_{\theta'}(x_i, Z_i=j)}{F_{ij}}$

$$\begin{aligned}
\frac{P_{\theta'}(x_i, Z_i = j)}{F_{ij}} &= \frac{P_{\theta'}(x_i, Z_i = j)}{P_{\theta'}(Z_I = j|x_i)} \\
&= \frac{P_{\theta'}(x_i, Z_i = j)}{\frac{P_{\theta'}(x_i, Z_i=j)}{P_{\theta'}(x_i)}} \\
&= P_{\theta'}(x_i)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}_{q_i}\left(\frac{P_{\theta'}(x_i, Z_i = j)}{F_{ij}}\right) &= \sum_{j=1}^{k} F_{ij} \frac{P_{\theta'}(x_i, Z_i = j)}{F_{ij}} \\
&= \sum_{j=1}^{k} P_{\theta'}(x_i, Z_i = j) \\
&= P_{\theta'}(x_i \cap (\cup_{j=1}^{k} Z_i = j)) \\
&= P_{\theta'}(x_i)
\end{aligned}
$$

Hence, all conditions are satisfied. $F_{ij} = P_{\theta'}(Z_i = j | x_i)$.

(e) & (f)

Before the derivation, we should note that $\pi_j^{(t+1)}$ and $\mu_j^{(t+1)}$ of the mixture of spherical Gaussians model and the mixture of diagonal Gaussians model are same. First, we derive $\pi_j^{(t+1)}$. The lower bound of function at $\theta^{(t)}$ is

$$
\begin{aligned}
Q(\theta^{(t)}, \theta) &= \sum_{i=1}^{n} \sum_{j=1}^{k} F_{ij}^{(t)} \log \left[\frac{P_\theta(x_i, Z_i = j)}{F_{ij}^{(t)}} \right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{k} F_{ij}^{(t)} \log \left[\frac{P_\theta(x_i | Z_i = j) P_\theta(Z_i = j)}{F_{ij}^{(t)}} \right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{k} F_{ij}^{(t)} \left(\log P_\theta(Z_i = j) + \log P_\theta(x_i | Z_i = j) - \log F_{ij}^{(t)} \right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{k} F_{ij}^{(t)} \left(\log \pi_j + \log P_\theta(x_i | Z_i = j) - \log F_{ij}^{(t)} \right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{k-1} F_{ij}^{(t)} \left(\log \pi_j + \log P_\theta(x_i | Z_i = j) - \log F_{ij}^{(t)} \right) \\
&\quad + \sum_{i=1}^{n} F_{ik}^{(t)} \left(\log \left(1 - \sum_{j=1}^{k-1} \pi_j \right) + \log P_\theta(x_i | Z_i = k) - \log F_{ik}^{(t)} \right)
\end{aligned}
$$

Then, for a specific $j \in \{1, 2, \cdots, k-1\}$, we calculate the partial derivative of function.

$$
\frac{\partial Q(\theta^{(t)}, \theta)}{\partial \pi_j} = \frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{\pi_j} - \frac{\sum_{i=1}^{n} F_{ik}^{(t)}}{\pi_k}
$$

In order to make this partial derivative equal 0, we need $\frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{\pi_j^{(t+1)}} = \frac{\sum_{i=1}^{n} F_{ik}^{(t)}}{\pi_k^{(t+1)}}$, that is, $\frac{\pi_j^{(t+1)}}{\pi_k^{(t+1)}} = \frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{\sum_{i=1}^{n} F_{ik}^{(t)}}$. Then, we can derive that

$$
\frac{1}{\pi_k^{(t+1)}} = \frac{\sum_{j=1}^{k-1} \pi_j^{(t+1)} + \pi_k^{(t+1)}}{\pi_k^{(t+1)}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} F_{ij}^{(t)}}{\sum_{i=1}^{n} F_{ik}^{(t)}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} F_{ij}^{(t)}}{\sum_{i=1}^{n} F_{ik}^{(t)}} = \frac{n}{\sum_{i=1}^{n} F_{ik}^{(t)}}
$$

Hence, $\pi_k^{(t+1)} = \frac{\sum_{i=1}^{n} F_{ik}^{(t)}}{n}$ and for $j \in \{1, 2, \cdots, k-1\}$, $\pi_j^{(t+1)} = \frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{\sum_{i=1}^{n} F_{ik}^{(t)}} \pi_k^{(t+1)} = \frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{n}$.
Furthermore, we verify that $\frac{\partial^2 Q(\theta^{(t)}, \theta)}{\partial^2 \pi_j} = -\frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{\pi_j^2} - \frac{\sum_{i=1}^{n} F_{ik}^{(t)}}{\pi_k^2} < 0$.
In conclusion, $\pi_j^{(t+1)} = \frac{\sum_{i=1}^{n} F_{ij}^{(t)}}{n}$.

Then, we derive $\mu_j^{(t+1)}$. From the above calculation, we have

$$Q(\theta^{(t)},\theta)=\sum_{i=1}^{n}\sum_{j=1}^{k}F_{ij}^{(t)}\left(\log\pi_j-\frac{d}{2}\log(2\pi)-\frac{1}{2}\log(|\Sigma_j|)-\frac{1}{2}(x_i-\mu_j)^\top\Sigma_j^{-1}(x_i-\mu_j)-\log F_{ij}^{(t)}\right)$$

Then, in order to obtain $\mu_j^{(t)}$, we need to solve the equation

$$\frac{\partial Q(\theta^{(t)},\theta)}{\partial\mu_j}=\sum_{i=1}^{n}F_{ij}^{(t)}\Sigma_j^{-1}(x_i-\mu_j)=\Sigma_j^{-1}\left(\sum_{i=1}^{n}F_{ij}^{(t)}x_i-\sum_{i=1}^{n}F_{ij}^{(t)}\mu_j\right)$$

Therefore,

$$\mu_j^{(t+1)}=\frac{\sum_{i=1}^{n}F_{ij}^{(t)}x_i}{\sum_{i=1}^{n}F_{ij}^{(t)}}$$

Next, we derive the $\Sigma_j^{(t+1)}$ for the mixture of spherical Guassains model. Because $\Sigma_j=\sigma_j^2 I_d$ for all $j=1,2,\cdots,k$, we only need to derive $(\sigma_j^2)^{(t+1)}$ and also can obtain $|\Sigma_j|=(\sigma_j^2)^d$, $\Sigma_j^{-1}=\frac{1}{\sigma_j^2}I_d$. Therefore,

$$\begin{aligned}Q(\theta^{(t)},\theta)&=\sum_{i=1}^{n}\sum_{j=1}^{k}F_{ij}^{(t)}\left(\log\pi_j-\frac{d}{2}\log(2\pi)-\frac{1}{2}\log(|\Sigma_j|)-\frac{1}{2}(x_i-\mu_j)^\top\Sigma_j^{-1}(x_i-\mu_j)-\log F_{ij}^{(t)}\right)\\&=\sum_{i=1}^{n}\sum_{j=1}^{k}F_{ij}^{(t)}\left(\log\pi_j-\frac{d}{2}\log(2\pi)-\frac{d}{2}\log(\sigma_j^2)-\frac{1}{2\sigma_j^2}(x_i-\mu_j)^\top(x_i-\mu_j)-\log F_{ij}^{(t)}\right)\end{aligned}$$

Then, in order to obtain $(\sigma_j^2)^{(t+1)}$, we need to solve the equation

$$\frac{\partial Q(\theta^{(t)},\theta)}{\partial\sigma_j^2}=\sum_{i=1}^{n}F_{ij}^{(t)}\left(-\frac{d}{2\sigma_j^2}+\frac{1}{2\sigma_j^4}(x_i-\mu_j)^\top(x_i-\mu_j)\right)=0$$

Because we have already obtained $\mu_j^{(t)}$, the solution is

$$(\sigma_j^2)^{(t+1)}=\frac{\sum_{i=1}^{n}F_{ij}^{(t)}(x_i-\mu_j)^\top(x_i-\mu_j)}{d\sum_{i=1}^{n}F_{ij}^{(t)}}$$

Last, we derive the $\Sigma_j^{(t+1)}$ for the mixture of diagonal Guassains model. Because $\Sigma_j=diag(\sigma_{j1}^2,\cdots,\sigma_{jd}^2)$ for $j=1,2,\cdots,k$, we only need to derive $(\sigma_{jm}^2)^{(t+1)}$ $(m=1,2,\cdots,d)$ and also can obtain $|\Sigma_j|=\prod_{m=1}^{d}\sigma_{jm}^2$, $(x_i-\mu_j)^\top\Sigma_j^{-1}(x_i-\mu_j)=\sum_{m=1}^{d}\frac{1}{\sigma_{jm}^2}(x_{im}-\mu_{jm})^2$. Therefore,

$$\begin{aligned}Q(\theta^{(t)},\theta)&=\sum_{i=1}^{n}\sum_{j=1}^{k}F_{ij}^{(t)}\left(\log\pi_j-\frac{d}{2}\log(2\pi)-\frac{1}{2}\log(|\Sigma_j|)-\frac{1}{2}(x_i-\mu_j)^\top\Sigma_j^{-1}(x_i-\mu_j)-\log F_{ij}^{(t)}\right)\\&=\sum_{i=1}^{n}\sum_{j=1}^{k}F_{ij}^{(t)}\left(\log\pi_j-\frac{d}{2}\log(2\pi)-\frac{1}{2}\sum_{m=1}^{d}\log(\sigma_{jm}^2)-\frac{1}{2}\sum_{m=1}^{d}\frac{1}{\sigma_{jm}^2}(x_{im}-\mu_{jm})^2-\log F_{ij}^{(t)}\right)\end{aligned}$$

Then, in order to obtain $(\sigma_{jm}^2)^{(t+1)}$, we need to solve the equation

$$\frac{\partial Q(\theta^{(t)},\theta)}{\partial\sigma_{jm}^2}=\sum_{i=1}^{n}F_{ij}^{(t)}\left(-\frac{1}{2\sigma_{jm}^2}+\frac{1}{2\sigma_{jm}^4}(x_{im}-\mu_{jm})^2\right)=0$$

Because we have already obtained $\mu_J^{(t)}$, the solution is

$$(\sigma_{jm}^2)^{(t+1)}=\frac{\sum_{i=1}^{n}F_{ij}^{(t)}(x_{im}-\mu_{jm})^2}{\sum_{i=1}^{n}F_{ij}^{(t)}}$$

Q3

As we all know, the max value of one pixel in a image is 255. It is too large to train the EM algorithm. Even though the EM algorithm can be trained, the accuracy in training data set only can reach 43%. It is so low that there is a huge space to raise the accuracy. In the end, we decided to normalization the original data to the interval of $[0 \sim 1]$. The operation of the normalization is only to divide 255.

The initial parameter of π_i is a vector which is $[0.2, 0.2, 0.2, 0.2, 0.2]$ for all first Spherical and Diagonal Gaussian Mixture Model, $[0.1, 0.2, 0.4, 0.2, 0.1]$ for all second Spherical and Diagonal Gaussian Mixture Model and $[0.2, 0.1, 0.4, 0.1, 0.2]$ for all third Spherical and Diagonal Gaussian Mixture Model.

The initial parameters μ_i and Σ_i are all initial with normal distribution random variable with different mean. For guarantee there are only positive value in the covariance matrix Σ_i, we used absolution operation to the random variables and add one to those random variables.

From the table we can know that the final log likelihood between the same Gaussian Mixture model is so tiny and it can be overlook. Because the final log likelihood is around same, the accuracy of training data set and testing data set are the same in the same Gaussian Mixture model.

	First Log-likelihood	End Log-likelihood	Accuracy of Train	Accuracy of Test
Spherical 1	-3481272.84	5952348.10115331	0.8057	0.8115
Spherical 2	-4683983.60	5952348.10079635	0.8057	0.8115
Spherical 3	-7513480.58	5952348.1008039	0.8057	0.8115
Diagonal 1	-3148557.53	6829014.62042487	0.7948	0.7989
Diagonal 2	-4464100.52	6829014.62006649	0.7948	0.7989
Diagonal 3	-9233210.69	6829014.62214698	0.7948	0.7989

Table 1: *The result of EM algorithm for different Gaussian Mixture Model.*