HW 3

ARNAB DEY
Student ID: 5563169
Email: dey00011@umn.edu

Solution 1

Definitions $\boldsymbol{z}^t$ is the vector of indicator variables, $\boldsymbol{z}^t = \{z_1^t, z_2^t, \ldots, z_K^t\}$, where $z_i^t = 1$ is $\boldsymbol{x}^t$ belongs to cluster G_i.
Total dataset is denoted as $X = \{\boldsymbol{x}^t\}_{t=1}^N$ and unobserved random variable dataset is $Z = \{\boldsymbol{z}^t\}_{t=1}^N$

Complete log-likelihood The complete log-likelihood function is given by:

$$\begin{aligned}
\mathcal{L}_c(\phi|X,Z) &= \ln p(X,Z|\phi) \\
&= \ln \prod_{t=1}^N p(\boldsymbol{x}^t, \boldsymbol{z}^t|\phi) \\
&= \sum_{t=1}^N \ln p(\boldsymbol{x}^t, \boldsymbol{z}^t|\phi) \\
&= \sum_{t=1}^N \ln \big(p(\boldsymbol{x}^t|\boldsymbol{z}^t, \phi) p(\boldsymbol{z}^t|\phi)\big) \\
&= \sum_{t=1}^N \left[\ln p(\boldsymbol{x}^t|\boldsymbol{z}^t, \phi) + \ln p(\boldsymbol{z}^t|\phi)\right] \\
&= \sum_{t=1}^N \left[\ln\left(\prod_{i=1}^K p_i(\boldsymbol{x}^t|\phi)^{z_i^t}\right) + \ln\left(\prod_{i=1}^K p(G_i)^{z_i^t}\right)\right] \\
&= \sum_{t=1}^N \left[\sum_{i=1}^K z_i^t \ln p(\boldsymbol{x}^t|\phi) + \sum_{i=1}^K z_i^t \ln p(G_i)\right] \\
&= \sum_{t=1}^N \sum_{i=1}^K z_i^t \left[\ln p(\boldsymbol{x}^t|\phi) + \ln \pi_i\right]
\end{aligned} \tag{1}$$

E-step: Here, we try to find the expectation of complete log-likelihood given the observed dataset and prior parameters ϕ. Thus,

$$\begin{aligned}
\mathcal{E}(\phi|\phi^l) &= \mathbf{E}[\mathcal{L}_c(\phi|X,Z)|X,\phi^l] \\
&= \mathbf{E}\left[\sum_{t=1}^N \sum_{i=1}^K z_i^t \left[\ln p(\boldsymbol{x}^t|\phi) + \ln \pi_i\right] |X, \phi^l\right] \\
&= \sum_{t=1}^N \sum_{i=1}^K \mathbf{E}[z_i^t|X,\phi^l] \left[\ln p(\boldsymbol{x}^t|\phi) + \ln \pi_i\right]
\end{aligned} \tag{2}$$

Now,

$$
\begin{aligned}
\mathbf{E}[z_i^t|X,\phi^l] &= p(z_i^t = 1|\boldsymbol{x}^t,\phi^l) \\
&= \frac{p(\boldsymbol{x}^t|z_i^t = 1,\phi^l)p(z_i^t = 1|\phi^l)}{\sum_{j=1}^{K} p(\boldsymbol{x}^t|z_j^t = 1,\phi^l)p(z_j^t = 1|\phi^l)} \\
&= \frac{p(\boldsymbol{x}^t|\phi^l)\pi_i}{\sum_{j=1}^{K} p(\boldsymbol{x}^t|\phi^l)\pi_j} \\
&= \frac{p(\boldsymbol{x}^t|G_i,\phi^l)p(G_i)}{\sum_{j=1}^{K} p(\boldsymbol{x}^t|G_j,\phi^l)p(G_j)} \\
&= p(G_i|\boldsymbol{x}^t,\phi^l) \\
&= \gamma(z_i^t)
\end{aligned}
\tag{3}
$$

Therefore, from Eq. (2), we try to formulate the maximization step as follows:

M-step: We try to maximize $\mathcal{E}(\phi|\phi^l)$:

$$
\phi^{l+1} = \arg\max_{\phi} \mathcal{E}(\phi|\phi^l)
$$

Now,

$$
\begin{aligned}
\mathcal{E}(\phi|\phi^l) &= \sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\left[\ln p(\boldsymbol{x}^t|\phi) + \ln \pi_i\right] \\
&= \sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln p(\boldsymbol{x}^t|\phi) + \sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln \pi_i
\end{aligned}
\tag{4}
$$

Maximization of priors, π_i: This is a constrained optimization with the constraint being $\sum_{i=1}^{K} \pi_i = 1$. Therefore, we use Lagrangian method to solve for π_i as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial \pi_i}\left[\sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln p(\boldsymbol{x}^t|\phi) + \sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln \pi_i - \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right)\right] = 0 \\
\implies &\frac{\sum_{t=1}^{N}\gamma(z_i^t)}{\lambda} = \pi_i
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\frac{\partial}{\partial \lambda}\left[\sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln p(\boldsymbol{x}^t|\phi) + \sum_{t=1}^{N}\sum_{i=1}^{K} \gamma(z_i^t)\ln \pi_i - \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right)\right] = 0 \\
\implies &\sum_{i=1}^{K}\pi_i = 1 \\
\implies &\lambda^* = N
\end{aligned}
$$

Therefore, the estimate of π_i is:

$$
\hat{\pi}_i = \frac{\sum_{t=1}^{N}\gamma(z_i^t)}{N}
$$

Maximization of parameters of the components: Here, it is given that,

$$
p(\boldsymbol{x}^t|\phi) = p(\boldsymbol{x}^t|\boldsymbol{\mu}_i,\boldsymbol{\sigma}_i) = \frac{1}{2\boldsymbol{\sigma}_i}e^{-\frac{|\boldsymbol{x}^t-\boldsymbol{\mu}_i|}{\boldsymbol{\sigma}_i}}
\tag{5}
$$

Therefore, from Eq. (4),

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\sigma}_i}\left[\sum_{t=1}^{N}\sum_{i=1}^{K}\gamma(z_i^t)\ln p(\boldsymbol{x}^t|\phi)+\sum_{t=1}^{N}\sum_{i=1}^{K}\gamma(z_i^t)\ln \pi_i\right]=0\\
\implies &\frac{\partial}{\partial \boldsymbol{\sigma}_i}\left[\sum_{t=1}^{N}\sum_{i=1}^{K}\gamma(z_i^t)\left(-\ln 2-\ln \boldsymbol{\sigma}_i-\frac{1}{\boldsymbol{\sigma}_i}|\boldsymbol{x}^t-\boldsymbol{\mu}_i|\right)\right]\\
\implies &\boldsymbol{S}_i=\frac{\sum_{t=1}^{N}\gamma(z_i^t)|\boldsymbol{x}^t-\boldsymbol{m}_i|}{\sum_{t=1}^{N}\gamma(z_i^t)},
\end{aligned} \tag{6}$$

where $\boldsymbol{m}_i$ is the MLE of $\boldsymbol{\mu}_i$ which is found as follows. To minimize Eq. (4) w.r.t. $\boldsymbol{\mu}_i$, we first binarize $\gamma(z_i^t)$ as $b_i^t = 1$ if $i = \arg\max_j \gamma(z_j^t)$. Then MLE of $\boldsymbol{\mu}_i$, $\boldsymbol{m}_i$ is given by,

$$\boldsymbol{m}_i = median\left(\{\boldsymbol{x}^t\}_{t=1}^{N_i}\right),$$

where, $N_i = \sum_{t=1}^{N} b_i^t$
Once we find the MLEs of the parameters, we prepare the next iteration:

$$\begin{aligned}
\boldsymbol{m}_i^{l+1} &= \boldsymbol{m}_i\\
\boldsymbol{S}_i^{l+1} &= \boldsymbol{S}_i\\
\gamma(z_i^t) &= \frac{\hat{\pi}_i 0.5\boldsymbol{S}_i^{-1}exp\left(-\frac{|\boldsymbol{x}^t-\boldsymbol{m}_i|}{\boldsymbol{S}_i}\right)}{\sum_{j=1}^{K}\hat{\pi}_j 0.5\boldsymbol{S}_j^{-1}exp\left(-\frac{|\boldsymbol{x}^t-\boldsymbol{m}_j|}{\boldsymbol{S}_j}\right)}
\end{aligned}$$

Solution 2.a

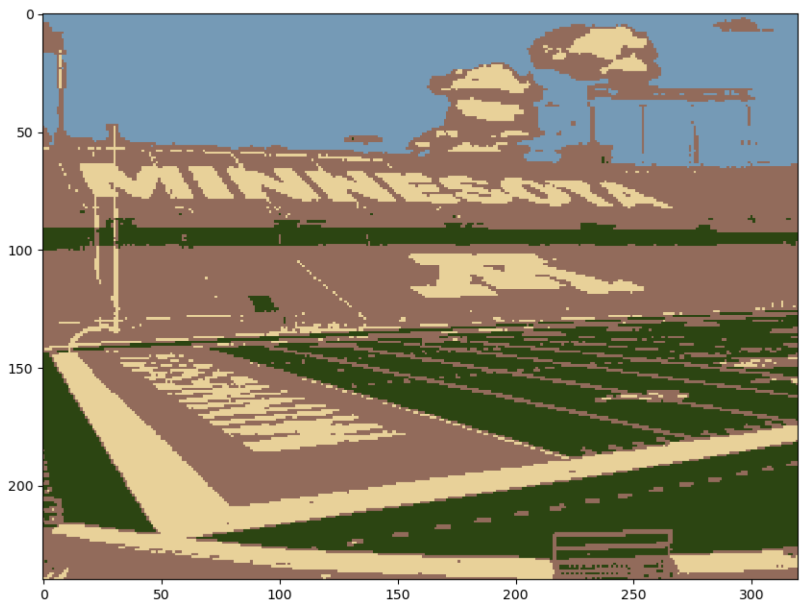

Figure 1: Q2.a: EM on 'stadium.jpg' with k=4 and 200 iterations

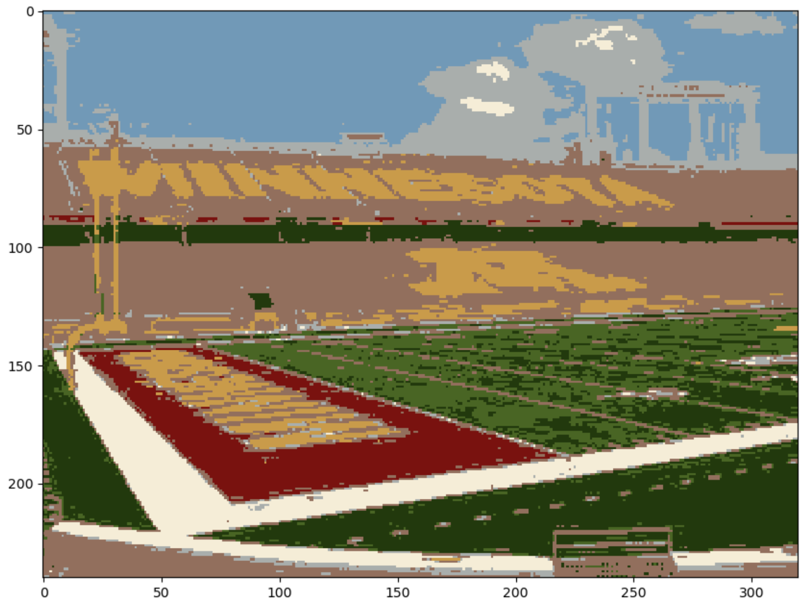

Figure 2: Q2.a: EM on 'stadium.jpg' with k=8 and 200 iterations

Figure 3: Q2.a: EM on 'stadium.jpg' with k=12 and 200 iterations

Solution 2.b

It can be seen from Fig. 4 that for each single value of k, expected value of complete log-likelihood gets maximized and after a certain iteration it does not increase any more which denotes convergence. Moreover, as the value of k increases, expected value of complete log-likelihood also increases which means higher the value of k, the image becomes closer to original one. Though it reduces compression ratio but gives higher quality picture.

Solution 2.c

Trying to run EM algorithm on 'goldy.jpg' raises runtime error while calculating $P(\boldsymbol{x}^t|\phi)$ as the covariance matrix becomes singular. Therefore, as the inverse of the covariance matrix cannot be calculated, we

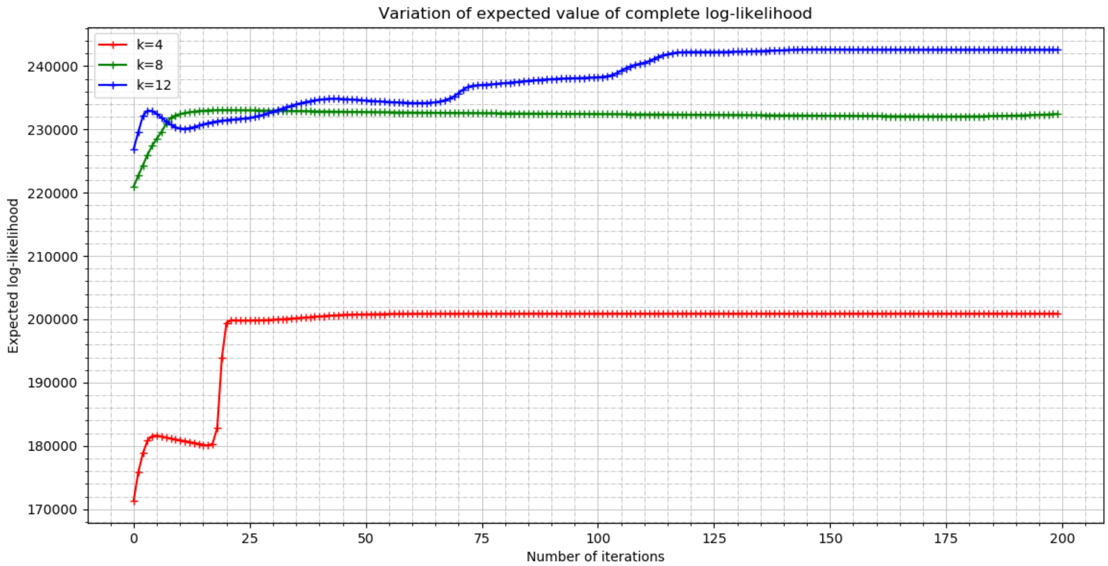

Figure 4: Q2.b: Expected value of log-likelihood on 'stadium.jpg' with different k

cannot calculate gaussian pdf value and cannot proceed further with E-step. Fig. (5) shows the result after

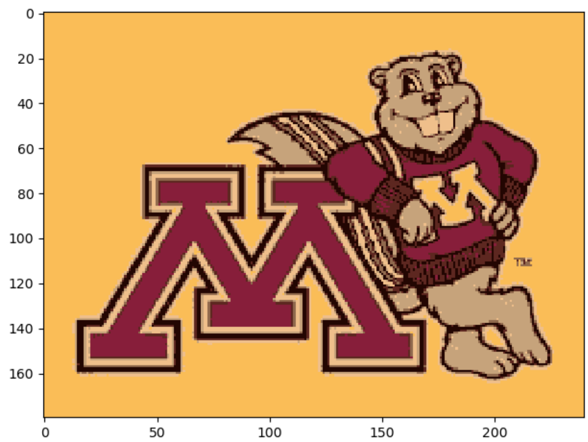

Figure 5: Q2.c: K-means on 'goldy.jpg'

compression of 'goldy.jpg' with k-means algorithm. The behavior of k-means and EM algorithm differs significantly as k-means does not use any parametric methods whereas in EM, the approach is probabilistic and depends on finding the parameters of underlying mixture of assumed probability density functions.

Solution 2.d

In the M-step, To find the estimate of $\mathbf{\Sigma}_i$ with regularization, first let us write the terms of expected complete log-likelihood function which depends on $\mathbf{\Sigma}_i$ as the other terms will eventually become 0 when we will take the derivative. We will also use the fact that

$$\boldsymbol{x}^T A\boldsymbol{x} = trace[\boldsymbol{x}^T A\boldsymbol{x}] = trace[\boldsymbol{x}\boldsymbol{x}^T A]$$

The expectation of complete log-likelihood function involving the terms that depend on $\boldsymbol{S}_i$ can be written as:

$$\begin{aligned}\mathcal{E}(\phi|\phi^l) &= \sum_{t=1}^{N}\sum_{i=1}^{K} h_i^t \ln p_i(\boldsymbol{x}^t|\phi^l) - \frac{\lambda}{2}\sum_{i=1}^{K}\sum_{j=1}^{d}\left(\mathbf{\Sigma}_i^{-1}\right)_{jj} \\ &= \sum_{t=1}^{N}\sum_{i=1}^{K} h_i^t \left[-\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}_i| - \frac{1}{2}\left(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1}\right)^T \mathbf{\Sigma}_i^{-1}\left(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1}\right)\right] - \frac{\lambda}{2}\sum_{i=1}^{K}\sum_{j=1}^{d}\left(\mathbf{\Sigma}_i^{-1}\right)_{jj}\end{aligned} \tag{7}$$

Removing the $\ln(2\pi)$ term as it is independent of $\mathbf{\Sigma}_i$, the above can be written as:

$$\begin{aligned}\mathcal{E}'(\phi|\phi^l) &= \sum_{t=1}^{N}\sum_{i=1}^{K}\left[-\frac{h_i^t}{2}\ln|\mathbf{\Sigma}_i| - \frac{h_i^t}{2}\left(trace\left((\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})^T\mathbf{\Sigma}_i^{-1}(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})\right)\right)\right] - \frac{\lambda}{2}\sum_{i=1}^{K}\sum_{j=1}^{d}\left(\mathbf{\Sigma}_i^{-1}\right)_{jj} \\ &= \sum_{t=1}^{N}\sum_{i=1}^{K}\left[-\frac{h_i^t}{2}\ln|\mathbf{\Sigma}_i| - \frac{h_i^t}{2}\left(trace\left((\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})^T\mathbf{\Sigma}_i^{-1}\right)\right)\right] - \frac{\lambda}{2}\sum_{i=1}^{K}\sum_{j=1}^{d}\left(\mathbf{\Sigma}_i^{-1}\right)_{jj}\end{aligned} \tag{8}$$

To find the estimate of $\mathbf{\Sigma}_i$, what we denote as $\boldsymbol{S}_i^{l+1}$, we can equivalently set the derivative of $\mathcal{E}'(\phi|\phi^l)$ w.r.t $\mathbf{\Sigma}_i^{-1}$ to 0, i.e.

$$\begin{aligned}&\frac{\partial\mathcal{E}'(\phi|\phi^l)}{\partial\mathbf{\Sigma}_i^{-1}} = 0 \\ &\implies 0 = \sum_{t=1}^{N} h_i^t\boldsymbol{S}_i^{l+1} - \sum_{t=1}^{N} h_i^t(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})^T - \frac{\lambda}{2}I \\ &\implies \boldsymbol{S}_i^{l+1} = \frac{\sum_{t=1}^{N}\left[h_i^t(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})(\boldsymbol{x}^t - \boldsymbol{m}_i^{l+1})^T\right] + \frac{\lambda}{2}I}{\sum_{t=1}^{N} h_i^t}\end{aligned} \tag{9}$$

Solution 2.e

Figure 6: Q2.e: EM with regularization $\lambda = 0.0001$ on 'goldy.jpg' and 200 iterations

When we add a regularization parameter, $\lambda = 0.0001$, the MLE of covariance matrix becomes non-singular as the regularization terms get added to the diagonal terms of covariance matrix. Therefore, we can see that the regularized EM can run the algorithm on 'goldy.jpg' and produces the Fig. 6.