Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2018

PA2

YOUR NAME: Guoqing Zhang November 5, 2018

- **Acknowledgments:** This template takes some materials from course CSE 547/Stat 548 of Washington University: https://courses.cs.washington.edu/courses/cse547/17sp/index.html.

  If you refer to other materials in your homework, please list here.

- **Collaborators:** I finish this homework by myself.

2.1.(a) Firstly, we need to know the log Maximum Likelihood Estimate:

$$
\begin{aligned}
&\log L(\mu_1, ..., \mu_k, \Sigma_1, ..., \Sigma_k, \phi_1, ..., \phi_k) \\
&= \log \prod_{i=1}^{m} p(x_i, y_i; \mu_1, ..., \mu_k, \Sigma_1, ..., \Sigma_k, \phi_1, ..., \phi_k) \\
&= \log \prod_{i=1}^{m} p(x_i|y_i; \mu_{y_i}, \Sigma_{y_i}) p(y_i; \phi_{y_i}) \\
&= \log \prod_{i=1}^{m} \prod_{j=1}^{k} \mathbf{1}\{y_i = j\} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_i-u_j)^T \Sigma^{-1}(x_i-u_j)} p(y_i = k; \phi_k) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{k} \mathbf{1}\{y_i = j\} (-\frac{1}{2}(x_i - u_j)^T \Sigma^{-1}(x_i - u_j) - \frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma_j| + \log p(y_i; \phi_{y_i}))
\end{aligned}
$$

If we want to find the Maximum, we need to get the derivative of Sigma. If we cut the useless parts,the function will be look like this:

$$
l = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{k} \mathbf{1}\{y_i = j\} (\log|\Sigma_j| - (x_i - u_j)^T \Sigma^{-1}(x_i - u_j))
$$

I need to tell some basic rules about derivative of matrix:

$$
\frac{\partial |A|}{\partial A} = |A|(A^{-1})^T \tag{1}
$$

$$
\frac{\partial A^{-1}}{\partial x} = A^{-1} \frac{\partial A}{\partial x} A^{-1} \tag{2}
$$

We could use the (1) to get the $\log|\Sigma_k|$'s derivative. Because of the SPD, we could get:

$$
\frac{\partial \log|\Sigma_j|}{\partial \Sigma_j} = (\Sigma_j^{-1})^T = \Sigma_j^{-1} \tag{3}
$$

Then, use the rule (2). Because the x is a scalar, so we need to separate the process.First let's try to find the derivative of $\Sigma_{k,(i,j)}$:

$$\begin{aligned}
\frac{\partial \Sigma_k^{-1}}{\partial \Sigma_{k,(i,j)}} &= \Sigma_k^{-1}\frac{\partial \Sigma_k}{\Sigma_{k,(i,j)}}\Sigma_k^{-1} \\
(x_i-u_j)^T\frac{\partial \Sigma_k^{-1}}{\partial \Sigma_{k,(i,j)}}(x_i-u_j) &= (x_i-u_j)^T\Sigma_k^{-1}\frac{\partial \Sigma_k}{\Sigma_{k,(i,j)}}\Sigma_k^{-1}(x_i-u_j)
\end{aligned}$$

We noticed that $(x_i-u_j)^T\Sigma_k^{-1} = (\Sigma_k^{-1}(x_i-u_j))^T$. And the matrix $\frac{\partial \Sigma_k^{-1}}{\partial \Sigma_{k,(i,j)}}$ will be like a n $\times$ n matrix with the exception that the value of the position(i,j) will be 1.

So we could get:

$$\begin{aligned}
(x_i-u_j)^T\frac{\partial \Sigma_k^{-1}}{\partial \Sigma_{k,(i,j)}}(x_i-u_j) &= (x_i-u_j)^T\Sigma_k^{-1}\frac{\partial \Sigma_k}{\Sigma_{k,(i,j)}}\Sigma_k^{-1}(x_i-u_j) \\
&= [(\Sigma_k^{-1}(x_i-u_j))(\Sigma_k^{-1}(x_i-u_j))^T]_{(i,j)}
\end{aligned}$$

So:

$$(x_i-u_j)^T\frac{\partial \Sigma_k^{-1}}{\partial \Sigma_{k,(i,j)}}(x_i-u_j) = (\Sigma_k^{-1}(x_i-u_j))(\Sigma_k^{-1}(x_i-u_j))^T \tag{4}$$

Now use (3) and (4),we could get:

$$\begin{aligned}
\frac{\partial l}{\partial \Sigma_j} &= \frac{1}{2}\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(\Sigma_j^{-1} - (\Sigma_k^{-1}(x_i-u_j))(\Sigma_k^{-1}(x_i-u_j))^T) \\
&= \frac{1}{2}\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(\Sigma_j^{-1} - \Sigma_j^{-1}(x_i-u_j)(x_i-u_j)^T\Sigma_j^{-1})
\end{aligned}$$

Because we want to let $\frac{\partial l}{\partial \Sigma_j} = \mathbf{0}$.

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(\Sigma_j^{-1} - \Sigma_j^{-1}(x_i-u_j)(x_i-u_j)^T\Sigma_j^{-1}) &= \mathbf{0} \\
\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(I - \Sigma_j^{-1}(x_i-u_j)(x_i-u_j)^T) &= \mathbf{0} \\
\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}I &= \Sigma_j^{-1}\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(x_i-u_j)(x_i-u_j)^T
\end{aligned}$$

So,for QDA,the Σ_j will be like this:

$$\Sigma_j = \frac{\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{m}\mathbf{1}\{y_i=j\}}$$

where $j = 1, 2$.

(b)The Programming assignment is attached in the zip file.