Name: Chee Kitt Win
Student Number: 45589140

a) i)

$$\log L(\theta) = \log b(x) - \log a(\theta) + c(\theta)^T t(x)$$

$$\Rightarrow \frac{\partial}{\partial\theta}\left(\log L(\theta)\right) = \frac{\partial}{\partial\theta}\left(c(\theta)^T t(x)\right) - \frac{\partial}{\partial\theta}\left(\log a(\theta)\right) \quad \text{— } 1$$

Since $E\left(\frac{\partial}{\partial\theta}\left(\log L(\theta)\right)\right) = 0$,

we have

$$E\left(\frac{\partial}{\partial\theta}\left(c(\theta)^T t(x)\right)\right) - E\left(\frac{\partial}{\partial\theta}\left(\log a(\theta)\right)\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}\left(\log a(\theta)\right) = E\left(\frac{\partial}{\partial\theta}\left(c(\theta)^T t(x)\right)\right) \quad \text{— } 2$$

2 $\rightarrow$ 1 yields

$$\frac{\partial}{\partial\theta}\left(\log L(\theta)\right) = \frac{\partial}{\partial\theta}\left(c(\theta)^T t(x)\right) - E\left(\frac{\partial}{\partial\theta}\left(c(\theta)^T t(x)\right)\right)$$

$$= \nabla_\theta\left(c(\theta)^T\right)\left(t(x)\right) - \nabla_\theta\left(c(\theta)^T\right)E\left(t(x)\right)$$

$$= \nabla_\theta\left(c(\theta)^T\right)\left(t(x) - E\left(t(x)\right)\right) \qquad \square$$

ii) The MLE is achieved at $\frac{\partial}{\partial\theta}\left(\log L(\theta)\right) = 0$

(For the regular exponential family, the natural param space is convex)

$$\frac{\partial}{\partial\theta}(\theta)\left(t(x) - E\left(t(x)\right)\right) = 0$$

$$\Rightarrow t(x) = E_{\hat{\theta}}\left(t(x)\right) \qquad \square$$

b) From part a) equation 2,

for $c(\theta)=\theta$,

$$\frac{\partial}{\partial\theta}\left(\log(a(\theta))\right) = E(t(x)) \quad \text{—} \ 4$$

$$L_c(\theta) = b(x)\left(\frac{1}{a(\theta)}\right)e^{\theta t(x)}$$

$$\Rightarrow \log L_c(\theta) = \log b(x) + \log\left(\frac{1}{a(\theta)}\right) + \theta t(x)$$

$$\Rightarrow E(\log L_c(\theta)) = E(\log b(x)) - E(\log(a(\theta))) + E(\theta t(x))$$

$$= E(\log(b(x))) - \log(a(\theta)) + \theta E(t(x))$$

$$\Rightarrow \text{Q function: } E_{\theta^{(k)}}(\log L_c(\theta)\,|\,y) = E_{\theta^{(k)}}(\log(b(x))\,|\,y) - \log(a(\theta)) + \theta E_{\theta^{(k)}}(t(x)\,|\,y)$$

Perform M-step:

$$\frac{\partial}{\partial\theta}\left(E_{\theta^{(k)}}(\log L_c(\theta)\,|\,y)\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}\left(\log(a(\theta))\right) = \frac{\partial}{\partial\theta}\left(\theta E_{\theta^{(k)}}(t(x)\,|\,y)\right) \quad \text{—} \ 3$$

Should go here

From 3 and 4, we have

$$\Rightarrow E(t(x)) = \frac{\partial}{\partial\theta}\left(\theta E_{\theta^{(k)}}(t(x)\,|\,y)\right)$$

$$\Rightarrow E_{\theta^{(k+1)}}(t(x)) = E_{\theta^{(k)}}(t(x)\,|\,y) \qquad \square$$

c) i)

$$\log L(\Psi) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{2} \pi_i f_i(w_j;\theta)\right)$$

$$\frac{\partial}{\partial \pi_1}\left(\log L(\Psi)\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial \pi_1}\left(\sum_{j=1}^{n} \log\left(\pi_1 f_1(w_j;\theta_1) + (1-\pi_1) f_2(w_j;\theta_2)\right)\right) = 0$$

$$\Rightarrow \sum_{j=1}^{n}\left(\frac{1}{\sum_{i=1}^{2} \pi_i f_i(w_j;\theta_i)}\right)\left(f_1(w_j;\theta_1) - f_2(w_j;\theta_2)\right) = 0$$

$$\Rightarrow \sum_{j=1}^{n}\left(\frac{f_1(w_j;\theta_1) - f_2(w_j;\theta_2)}{f(w_j;\Psi)}\right) = 0$$

mult by π_1

$$\Rightarrow \sum_{j=1}^{n}\left(\frac{\pi_1 f_1(w_j;\theta) - \pi_1 f_2(w_j;\theta_2)}{f(w_j;\Psi)}\right) = 0$$

$$\Rightarrow \sum_{j=1}^{n}\left(T_1(w_j;\Psi) - \frac{\pi_1}{1-\pi_1} T_2(w_j;\Psi)\right) = 0$$

$$\Rightarrow (1-\pi_1)\sum_{j=1}^{n}\left(T_1(w_j;\Psi)\right) = \pi_1 \sum_{j=1}^{n}\left(T_2(w_j;\Psi)\right)$$

$$\Rightarrow \sum_{j=1}^{n}\left(T_1(w_j;\Psi)\right) = \pi_1 \sum_{j=1}^{n}\left(T_1(w_j;\Psi) + T_2(w_j;\Psi)\right)$$

$$= \pi_1 \sum_{j=1}^{n}(1)$$

$$= n\pi_1$$

$$\Rightarrow \hat{\pi}_1 = \frac{1}{n}\sum_{j=1}^{n}\left(T_1(w_j;\Psi)\right) \quad — (5)$$

$$\frac{\partial}{\partial \theta_i}\left(\log L(\Psi)\right) = 0 \Rightarrow \sum_{j=1}^{n}\left(\frac{\pi_i \frac{\partial}{\partial \theta_i}\left(f_i(w_j;\theta_i)\right)}{f(w_j;\Psi)}\right) = 0$$

$$\Rightarrow \sum_{j=1}^{n}\left(\frac{\pi_i f_i(w_j;\theta_i)\frac{\partial}{\partial \theta_i}\left(f(w_j;\theta_i)\right)}{f(w_j;\Psi) f_i(w_j;\theta_i)}\right) = 0$$

$$\Rightarrow \sum_{j=1}^{n}\left(T_i(w_j;\Psi)\frac{\partial}{\partial \theta_i}\left(\log f(w_j;\theta_i)\right)\right) = 0 \quad — (6)$$

~~of indicator variables where $Z_{ij} = (Z_j)_i$ is one if w_j comes from the ith component of the mixture and zero otherwise~~

~~Notation: $i \in \{1, 2\}$, $j \in \{1, 2, \ldots, n\}$~~

~~We then~~

ii) We first substitute an initial value for $\Psi = (\pi_1, \theta_1^T, \theta_2^T)$ into the RHS of (5) and (6) in part i). This yields a new value of Ψ. Repeat this prosess till convergence. (Until the difference between the LHS and RHS of (5) and (6) changes by an arbitrarily small amount)

d) i) We can write $f(w_j, z_j; \Psi) = P(Z_j = z_j; \Psi) \cdot f(W_j = w_j | Z_j = z_j; \Psi)$

$$= \prod_{i=1}^{g} \pi_i^{z_{ij}} \phi(w_j; \mu_i, \Sigma_i)^{z_{ij}}$$

$$\Rightarrow L_c(\Psi) = \prod_{j=1}^{n} \prod_{i=1}^{g} \pi_i^{z_{ij}} \phi(w_j; \mu_i, \Sigma_i)^{z_{ij}}$$

$$= \prod_{j=1}^{n} \prod_{i=1}^{g} \left[\pi_i^{z_{ij}} \left(\frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (w_j - \mu_i)^T \Sigma_i^{-1} (w_j - \mu_i) \right) \right)^{z_{ij}} \right]$$

ii) $Q(\Psi; \Psi^{(k)}) = E_{\Psi^{(k)}} (\log L_c(\Psi) | y)$

$$= E_{\Psi^{(k)}} \left[\sum_{j=1}^{n} \sum_{i=1}^{g} \left(z_{ij} \log \pi_i + z_{ij} \log \phi(w_j; \mu_i, \Sigma_i) \right) \Big| y \right]$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{g} \left[\left(E_{\Psi^{(k)}} (z_{ij} | w_j) \right) \left(\log \pi_i + \log \phi(w_j; \mu_i, \Sigma_i) \right) \right] \quad (7)$$

$$E_{\Psi^{(k)}}(Z_{ij}|W_j) = 0 \cdot P(Z_{ij}=0 \mid W_j = w_j) + 1 \cdot P(Z_{ij}=1 \mid W_j = w_j)$$

$$= \frac{f(Z_{ij}=1, W_j = w_j)}{f(W_j)}$$

$$= \frac{\pi_i^{(k)} \phi_i(w_j; \mu_i^{(k)}, \Sigma_i^{(k)})}{\sum_{i=1}^{g} \pi_i^{(k)} \phi_i(w_j; \mu_i^{(k)}, \Sigma_i^{(k)})}$$

$$= \tau_i(w_j; \Psi^{(k)}) \quad \text{— } 8$$

8 -> 7, we have

$$Q(\Psi; \Psi^{(k)}) = \sum_{j=1}^{n} \sum_{i=1}^{g} \tau_i(w_j; \Psi^{(k)}) \left(\log \pi_i + \log \phi_i(w_j; \mu_i, \Sigma_i)\right)$$

iii) From the notes on finite normal mixtures, we know that, assuming z is observable, we have the following MLE's

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} z_{ij} w_j}{\sum_{j=1}^{n} z_{ij}}, \quad \hat{\Sigma}_i = \frac{\sum_{j=1}^{n} z_{ij}(w_j - \mu_i)(w_j - \mu_i)^T}{\sum_{j=1}^{n} z_{ij}}, \quad \hat{\pi}_i = \frac{\sum_{j=1}^{n} z_{ij}}{n}$$

From 7 in part ii), we see that z_{ij} has simply been replaced by $E_{\Psi^{(k)}}(Z_{ij}|w_j) = \tau_i(w_j; \Psi^{(k)})$ in the Q function. Therefore, we can replace z_i by $\tau_i(w_j; \Psi^{(k)})$ in the above MLE's to yield the following update equations for the $(k+1)$th iteration.

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)}) w_j}{\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})} \qquad \Sigma_i^{(k+1)} = \frac{\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})(w_j - \mu_i^{(k+1)})(w_j - \mu_i^{(k+1)})^T}{\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})}$$

$$\pi_i^{(k+1)} = \frac{\sum_{j=1}^{n} \tau(w_j; \Psi^{(k)})}{n}$$

e) i) Using $E(f(x)) = E_{\theta^{(k)}}(f(x)|y)$, we have,

for $\sum_{j=1}^{n} z_{ij}$,

$$E\left(\sum_{j=1}^{n} z_{ij}\right) = E_{\theta^{(k)}}\left(\sum_{j=1}^{n} z_{ij} \Big| y\right)$$

$$\Rightarrow \sum_{j=1}^{n} E(z_{ij}) = \sum_{j=1}^{n} E_{\theta^{(k)}}(z_{ij}|w_j)$$

from (8) in d) ii)

$$\Rightarrow \sum_{j=1}^{n} (1)\cdot P(z_{ij}=1) = \sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \sum_{j=1}^{n} \pi_i = \sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \pi_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)}) \quad \text{——} \ (9)$$

for $\sum_{j=1}^{n}(z_{ij}w_j)$,

$$E\left(\sum_{j=1}^{n}(z_{ij}w_j)\right) = E_{\theta^{(k)}}\left(\sum_{j=1}^{n}(z_{ij}w_j) \Big| y\right)$$

$$\Rightarrow \sum_{j=1}^{n} E(z_{ij}w_j) = \sum_{j=1}^{n} E_{\theta^{(k)}}(z_{ij}w_j | w_j)$$

$$\Rightarrow \sum_{j=1}^{n} E(E(z_{ij}w_j|w_j)) = \sum_{j=1}^{n} w_j E_{\theta^{(k)}}(z_{ij}|w_j)$$

$$\Rightarrow \sum_{j=1}^{n} E(w_j \tau_i(w_j; \Psi)) = \sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \sum_{j=1}^{n} \int_{w_j} w_j \tau_i(w_j; \Psi) f(w_j)\, dw_j = \sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \sum_{j=1}^{n} \int_{w} (w_j)\left(\frac{\pi_i \phi_i(w_j; \mu_i, \Sigma_i)}{f(w_j)}\right) f(w_j)\, dw_j = \sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \pi_i \sum_{j=1}^{n} \int_{w} w_j \phi_i(w_j; \mu_i, \Sigma_i)\, dw_j = \sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})$$

$$\Rightarrow \pi_i n \mu_i = \sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})$$

Sub (9)

$$\Rightarrow \mu_i^{(k+1)} = \frac{\sum_{j=1}^{n} w_j \tau_i(w_j; \Psi^{(k)})}{\sum_{j=1}^{n} \tau_i(w_j; \Psi^{(k)})}$$

for $\sum_{j=1}^{n} Z_{ij} W_j W_j^T$,

$$E\left(\sum_{j=1}^{n} Z_{ij} W_j W_j^T\right) = E_{\theta^{(k)}}\left(\sum_{j=1}^{n} Z_{ij} W_j W_j^T \middle| y\right)$$

RHS: $E_{\theta^{(k)}}\left(\sum_{j=1}^{n} Z_{ij} W_j W_j^T \middle| y\right) = \sum_{j=1}^{n} E_{\theta^{(k)}}\left(Z_{ij} W_j W_j^T / W_j\right)$

$$= \sum_{j=1}^{n} W_j W_j^T E_{\theta^{(k)}}\left(Z_{ij} | W_j\right)$$

Again from (8)

$$= \sum_{j=1}^{n} W_j W_j^T \tau_i\left(W_j ; \Psi^{(k)}\right) \quad \text{— (10)}$$

LHS: $E\left(\sum_{j=1}^{n} Z_{ij} W_j W_j^T\right) = \sum_{j=1}^{n} E\left(E\left(Z_{ij} W_j W_j^T / W_j\right)\right)$

From RHS

$$= \sum_{j=1}^{n} E\left(W_j W_j^T \tau_i\left(W_j ; \Psi\right)\right)$$

$$= \sum_{j=1}^{n} \int_W W_j W_j^T \left(\frac{\pi_i \phi_i(W_j ; \mu_i, \Sigma_i)}{f(W_j ; \Psi)}\right) f(W_j ; \Psi) \, dw_j$$

$$= \pi_i \sum_{j=1}^{n} \int_W W_j W_j^T \phi_i(W_j ; \mu_i, \Sigma_i) \, dw_j$$

$$= \pi_i \sum_{j=1}^{n} E_{\phi_i}\left(W_j W_j^T\right)$$

$$= n \pi_i E_{\phi_i}\left(W_j W_j^T\right) \quad \text{— (11)}$$

Equating LHS (11) and RHS (10),

$$E_{\phi_i}\left(W_j W_j^T\right) = \frac{\sum_{j=1}^{n} W_j W_j^T \tau_i(W_j ; \Psi^{(k)})}{n \pi_i}$$

$$\Rightarrow E_{\phi_i}\left(W_j W_j^T\right) - \mu_i \mu_i^T = \frac{\sum_{j=1}^{n} W_j W_j^T \tau_i(W_j ; \Psi^{(k)})}{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)})} - \mu_i \mu_i^T$$

$$\Rightarrow \Sigma_i^{(k+1)} = \frac{\sum_{j=1}^{n} W_j W_j^T \tau_i(W_j ; \Psi^{(k)}) - \mu_i \mu_i^T \sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)})}{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)})}$$

$$= \frac{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)}) (W_j W_j^T - \mu_i \mu_i^T)}{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)})}$$

$$= \frac{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)}) (W_j - \mu_i^{(k+1)})(W_j - \mu_i^{(k+1)})^T}{\sum_{j=1}^{n} \tau_i(W_j ; \Psi^{(k)})} \quad \#$$