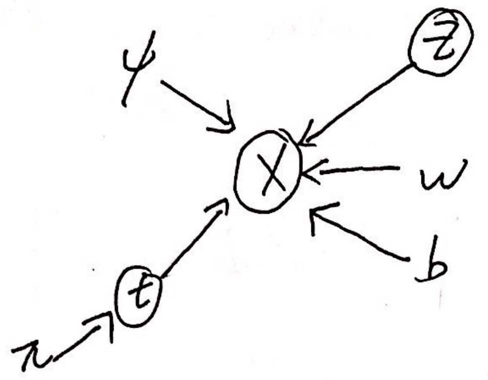
# COMP 9418 ASSIGMENT 2

ZID: Z5127440
Name: YIFAN ZHAO

a)




b)

For $\psi$:

$\because \psi \in R^{D \times D}$ is a diagonal covariance matrix,

so the number of parameter for $\psi$ is $\underline{D}$.

For $W_k$:

$\because W_k \in R^{D \times Q}$, $\forall k$, $W_k$ is a $D \times Q$ matrix,

So the number of parameters for $W_k$ is $\underline{k \times D \times Q}$.

For $b_k$:

$\because b_k \in R^D$, $\forall k$, $b_k$ is a D-dimensional vector,

So the number of parameters for $b_k$ is $\underline{k \times D}$.

For $\pi$,

$\because \sum \pi = 1$, if we know $(k-1)$ values, we can calculate the $k^{th}$ value, so the number of parameters is $\underline{k-1}$.

In conclusion,

the number of parameters we are required is

$$\underline{D + kDQ + kD + k-1}.$$

c)

i) (1) We have:

$$\gamma_{nk} = E_{p(t_{nk} | x_n \theta^{old})}[t_{nk}] = 0 \times p(t_{nk}=0 | x_n \theta^{old}) + 1 \times p(t_{nk}=1 | x_n, \theta^{old})$$

$$= p(t_{nk}=1 | x_n, \theta^{old})$$

$$\boxed{\begin{aligned} \text{Theory: } \quad & p(x,y) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x) \\ & \Rightarrow p(x|y) = \frac{p(y|x) \cdot p(x)}{p(y)} \quad (*) \end{aligned}}$$

Let $(t_{nk}=1) = x$, $x_n = y$ in $(*)$, we get:

$$\gamma_{nk} = p(t_{nk}=1 | x, \theta^{old}) = \frac{p(x_n | t_{nk}=1, \theta^{old}) \cdot p(t_{nk}=1)}{p(x)}$$

Then $p(t_{nk}=1 | x, \theta^{old}) \propto p(x_n | t_{nk}=1, \theta^{old}) \cdot p(t_{nk}=1)$

Now, we add $z$ into the expression.

$$p(x_n | t_{nk}=1, z, \theta^{old}) \cdot p(z_n | t_{nk}=1, \theta) \cdot p(t_{nk}=1)$$

From (a), we now $z$ and $t_{nk}$ are indopent.

$$\therefore p(z_n | t_{nk}=1, \theta^{old}) = p(z_n | \theta)$$

$\therefore p(X_n | t_{nk}=1, Z_n, \theta^{old}) \, p(Z_n) \, p(t_{nk}=1)$

$= N(X_n | W_k z + b_k, \psi) \cdot N(Z_n | 0, I) \cdot \pi_k$

Hence, $r_{nk} = \pi_k N(X_n - b_k, W_k W_k^T + \psi)$

(2) We have $m_{nk} \overset{def}{=} E_{p(Z_n | t_{nk}=1, X_n, \theta^{old})} [Z_n]$

$\because$ This is a linear projection, $E(Z|X) = \beta X$

where $\beta_k = W_k^T (\psi + W_k W_k^T)^{-1}$ ,

Hence, $E[p(Z_n | t_{nk}=1, X_n, \theta^{old})] = \beta_k (X_n - b_k)$

$\therefore m_{nk} = W_k^T (\psi + W_k W_k^T)^{-1} (X_n - b_k)$

(3) We have $C_{nk} \overset{def}{=} S_{nk} - m_{nk} m_{nk}^T$

$\therefore S_{nk} = C_{nk} + m_{nk} m_{nk}^T$

$\because C_{nk}$ is covariance of conditional posterior over local hidden factor $Z_n$.

$p(Z_n | t_{nk}=1, X_n, \theta^{old}) = N(Z_n | m_{nk}, C_{nk})$

Hence, $S_{nk} = I - W_k^T (\psi + W_k W_k^T)^{-1} W_k + W_k^T (\psi + W_k W_k^T)^{-1} (X_n - b_k)(X_n - b_k)^T (W_k^T (\psi + W_k W_k^T)^{-1})^T$

(ii) $\because \tilde{m}_{nk} \stackrel{def}{=} E_{p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old})}[\tilde{z}_n]$

$$\tilde{z}_n \stackrel{def}{=} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \\ 1 \end{bmatrix}$$

$$p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old}) = p(\begin{bmatrix} z \\ 1 \end{bmatrix} | t_{nk}=1, x_n, \theta^{old})$$

$$\therefore E(\tilde{z}_n | x_n, t_{nk}) = \begin{bmatrix} m_{nk} \\ 1 \end{bmatrix}$$

$$\therefore \tilde{m}_{nk} = \begin{bmatrix} m_{nk} \\ 1 \end{bmatrix}$$

$$\because \tilde{S}_{nk} \stackrel{def}{=} E_{p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old})}[\tilde{z}_n \cdot \tilde{z}_n^T]$$

$$\tilde{z}_n \tilde{z}_n^T = \begin{bmatrix} z_n z_n^T & z_n \\ z_n^T & 1 \end{bmatrix}$$

$$\text{Hence } E[\tilde{z}_n \tilde{z}_n^T | x_n, t_{nk}] = \begin{pmatrix} E[z_n z_n^T | x_n, t_{nk}] & E[z_n | x_n, t_{nk}] \\ E[z_n | x_n, t_{nk}]^T & 1 \end{pmatrix}$$

$$= \begin{bmatrix} S_{nk} & m_{nk} \\ m_{nk}^T & 1 \end{bmatrix}$$

(iii) (1) $E_{p(\tilde{z}_n; t_{nk} | x_n, \theta^{old})}[t_{nk} \tilde{z}_n] \Rightarrow p(\tilde{z}_n, t_{nk} | x_n, \theta^{old})$

$= p(t_{nk} | x_n, \tilde{z}_n, \theta^{old}) \cdot p(\tilde{z}_n | x_n, \theta^{old})$

because of formula (6): $\gamma_{nk} = E_{p(t_{nk} | x_n, \theta^{old})}[t_{nk}]$

$\because$ $t_{nk}$ and $z_n$ are independent

$\therefore E_{p(t_{nk} | x_n, \theta^{old})}[t_{nk}] = E_{p(t_{nk} | x_n, \tilde{z}_n, \theta^{old})} = \gamma_{nk}$

$\because$ Formula (11) $\tilde{m}_{nk} = E_{p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old})}[\tilde{z}_n]$

$E_{p(\tilde{z}_n, t_{nk} | x_n, \theta^{old})} = E[p(t_{nk} | x_n, \tilde{z}_n, \theta^{old})] \cdot E[p(\tilde{z}_n | x_n, \theta^{old})]$

$\because E_{p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old})}[\tilde{z}_n] = E_{p(\tilde{z}_n | x_n, \theta^{old})} = \tilde{m}_{nk}$

$\therefore E_{p(\tilde{z}_n, t_{nk} | x_n, \theta^{old})} = \gamma_{nk} \cdot \tilde{m}_{nk}$

(2) $E_{p(\tilde{z}_n, t_{nk} | x_n, \theta^{old})}[t_{nk} \tilde{z}_n \tilde{z}_n^T] = E[p(t_{nk} | x_n, \tilde{z}_n, \theta^{old})] \cdot E[p(\tilde{z}_n | x_n, \theta^{old})]$

because of: $p(\tilde{z}_n, t_{nk} | x_n, \theta^{old}) = p(t_{nk} | x_n, \tilde{z}_n, \theta^{old}) \cdot p(\tilde{z}_n | x_n, \theta^{old})$

The same reason: From formula (6): $\gamma_{nk} = E_{p(t_{nk} | x, \theta^{old})}[t_{nk}]$

$= E_{p(t_{nk} | x_n, \tilde{z}_n, \theta^{old})}$

From formula (11): $\tilde{S}_{nk} = E_{p(\tilde{z}_n | t_{nk}=1, x_n, \theta^{old})}[\tilde{z}_n \tilde{z}_n^T]$

$= E_{p(\tilde{z}_n | x_n, \theta^{old})}[\tilde{z}_n \cdot \tilde{z}_n^T]$

$\therefore E_{p(\tilde{z}_n, t_{nk} | x_n, \theta^{old})}[t_{nk} \tilde{z}_n \tilde{z}_n^T] = \gamma_{nk} \tilde{S}_{nk}$

(d) $\because$ $Q(\theta, \theta^{old}) \overset{def}{=} E_{p(Z,T|X,\theta^{old})}[\log P(Z,T,X|\theta)]$

$P(Z,T,X|\theta) = P(X|Z,T,\theta)\, P(Z,T|\theta)$

From (a), we have $Z$ and $T$ are independent

$\therefore P(Z,T|\theta) = P(Z|\theta) \cdot P(T|\theta)$

$\therefore P(Z,T,X|\theta) = P(X|Z,T,\theta) \cdot P(Z|\theta) \cdot P(T|\theta)$

we use formula (5), (3), (2)

(5): $P(X|Z,T,\theta) = \prod_1^k N(X|W_k Z + b_k, \psi)^{t_k}$, for $X_n$ in $X \in \{X_n\}_1^N$

(3): $P(Z|\theta) = P(Z) = N(0, I)$, for $Z_n \in \{Z_n\}_1^N$

(2): $P(T|\theta) = \prod_1^k \pi_k^{t_k}$, for each $t_n$ in $T \in \{t_n\}_1^N$

Hence, $P(Z,T,X|\theta) = \prod_1^N \prod_1^k N(X_n | W_k Z_n + b_k, \psi)^{t_k} \cdot N(0,I) \cdot \pi_k^{t_k}$

Then, the expected log likelihood is

$$Q = E\left[\log \prod_1^N \prod_1^k \left\{(2\pi)^{\frac{D}{2}} |\psi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[X_n - b_k - W_k Z_n]^T \psi^{-1} [X_n - b_k - W_k Z_n]\right\}\right\}^{t_k}\right]$$

$\because \tilde{Z} = \begin{bmatrix} Z \\ 1 \end{bmatrix}$ $\quad \tilde{W}_k = [W_k, b_k]$

$$\therefore Q(\theta, \theta^{old}) = E\left[\log \prod_1^N \prod_1^k \left\{(2\pi)^{\frac{D}{2}} |\psi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[X_n - \tilde{W}_k \tilde{Z}_n]^T \psi^{-1} [X_n - \tilde{W}_k \tilde{Z}_n]\right\}\right\}^{t_k}\right]$$

$$= C - \frac{N}{2}\log(\psi) - \sum_n \sum_k \frac{1}{2} r_{nk} X_n^T \psi^{-1} X_n - r_{nk} X_n^T \psi^{-1} \tilde{W}_k \tilde{m}_{nk} + \frac{1}{2} r_{nk} \cdot \mathrm{tr}[\tilde{W}_k^T \psi^{-1} \tilde{W}_k S_{nk}]$$

where $C$ is a constant

(e)

(1) $\because \tilde{W}_k = [W_k, b_k] \quad \tilde{z} = \begin{bmatrix} z \\ 1 \end{bmatrix}$

$$\frac{dQ}{d\tilde{W}_k} = -\sum_n r_{nk} \psi^{-1} x_n \tilde{m}_{nk}^T + r_{nk} \psi^{-1} \tilde{W}_k^{new} \tilde{S}_{nk} = 0$$

Hence $[W_k^{new}, b_k^{new}] = \tilde{W}_k^{new} = (\sum_n r_{nk} x_n \tilde{m}_{nk}^T)(\sum_l r_{lk} \tilde{S}_{lk})^{-1}$

where $\tilde{m}_{nk} = \begin{bmatrix} m_{nk} \\ 1 \end{bmatrix} \quad \tilde{S}_{nk} = \begin{bmatrix} S_{nk} & m_{nk} \\ m_{nk}^T & 1 \end{bmatrix}$

(2) To re-estimate the mixing proportions we use the defination.

$$\pi_k = p(t_k) = \int p(t_k|x) p(x) dx$$

$\because r_{nk} = p(t_k|x_n)$, we use empirical distribution

$$\pi_k^{new} = \frac{1}{N} \sum_1^N r_{nk}$$

(3) $$\frac{dQ}{d\psi^{-1}} = \frac{N}{2}\psi^{new} - \sum_{nk} \frac{1}{2} r_{nk} x_n x_n^T - r_{nk} \tilde{W}_k^{new} \tilde{m}_{nk} x_n^T$$

$$+ \frac{1}{2} r_{nk} \tilde{W}_k^{new} \tilde{S}_{nk} \tilde{W}_k^{new T} = 0$$

Then $\psi^{new} = \frac{1}{N} \mathrm{diag}\{ \sum_{nk} r_{nk} (x_n - \tilde{W}_k^{new} \tilde{m}_{nk}) x_n^T \}$