Q1.

a) With single link, once we cluster certain points, the next point will be associated to the cluster based on the minimum distance.

From the given table, the distance between x_1 and x_2 is minimum. (i.e 0.12). Hence proceeding by considering them as cluster

Step 1:

	x_1x_2	x_3	x_4	x_5	x_6
x_1x_2	0				
x_3	0.25	0			
x_4	0.16	minimum $\boxed{0.14}$	0		
x_5	0.28	0.70	0.45	0	
x_6	0.34	0.93	0.20	0.67	0

Step 2:

	x_1x_2	x_3x_4	x_5	x_6
x_1x_2	0			
x_3x_4	$\boxed{0.16}$	0		
x_5	0.28	0.45	0	
x_6	0.34	0.20	0.67	0

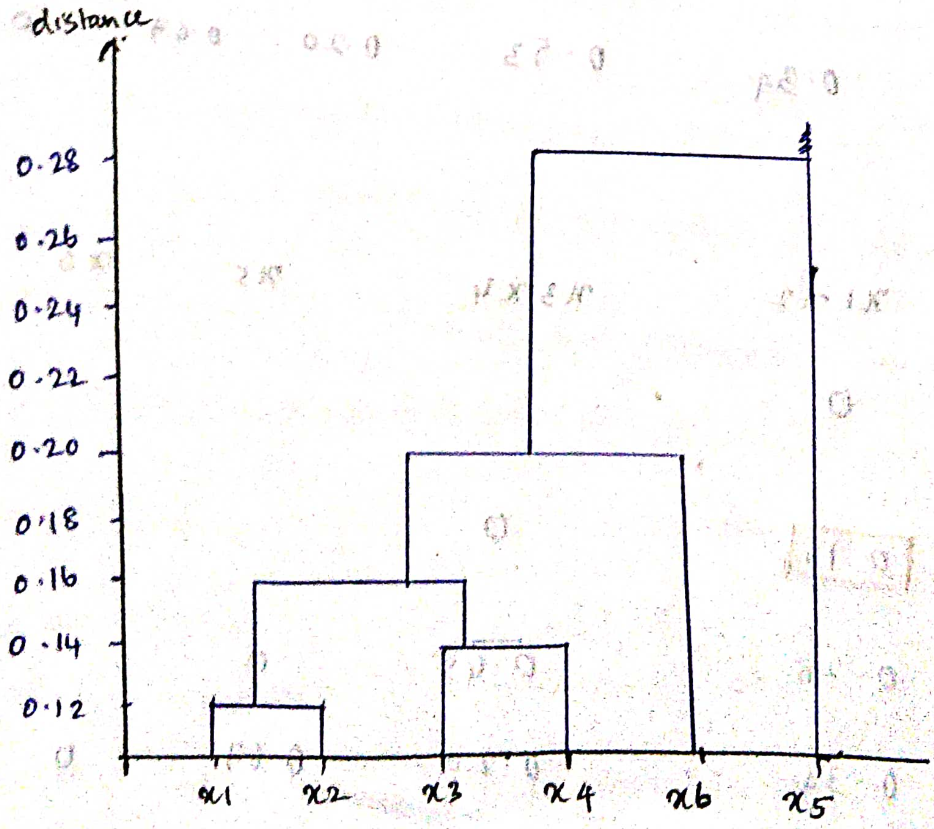
Step 3:

	$x_1 x_2 x_3 x_4$	x_5	x_6
$x_1 x_2 x_3 x_4$	0		
x_5	0.28	0	
x_6	$\boxed{0.20}$	0.67	0

Step 4:

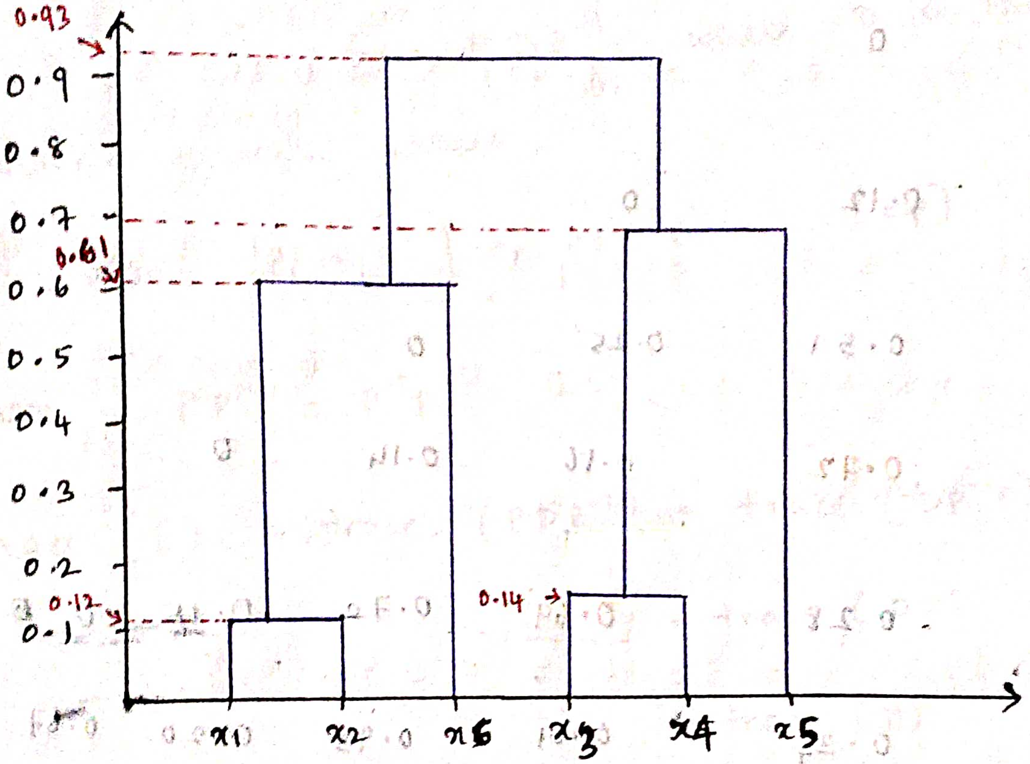
	$x_1 x_2 x_3 x_4 x_6$	x_5
$x_1 x_2 x_3 x_4 x_6$	0	
x_5	$\boxed{0.28}$	0

Hence the dendrogram is,



b) In case of complete link, once we cluster points, the next point will be associated with the cluster based on the <u>maximum</u> <u>distance</u>.

Proceeding with the steps similar to previous question, we get the dendrogram as,



1c)

W.e could change values as a result of binary decision we are taking based on minimum and maximum distance.

For example, in single link, once we cluster x_1x_2 the distance metric between x_1x_2 and x_3 is decided based on $\min(x_1x_3, x_2x_3)$.

ex: $\min(0.51, 0.25)$, here as we choose 0.25, the 0.51 can be anything other than 0.51 (ie > 0.25) such that it wont affect the dendrogram with complete link too. Hence by trying out various values, the modified table is,

	$x1$	$x1$	$x3$	$x4$	$x5$	$x6$
$x1$	0					
$x2$	0.12	0				
$x3$	0.51	0.25	0			
$x4$	$\boxed{0.72}$	0.16	0.14	0		
$x5$	0.28	$\boxed{0.69}$	0.70	0.45	0	
$x6$	0.34	0.61	0.93	0.20	0.67	0

2 a)

Since Σ is a covariance matrix, the sum of the diagonal elements is the trace.

i.e $\sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp} = \text{trace}(\Sigma)$

We can write the covariance matrix,

$\Sigma = PDP^{-1}$ where D is the diagonal matrix of eigen values.

also, $P_1 \Rightarrow [e_1, e_2, e_3, \ldots, e_p]$

here $PP^{-1} = P^{-1}P = I$

$$\text{trace}(\Sigma) = \text{trace}(PDP^{-1}) = \text{trace}(DP^{-1}P)$$
$$= \text{trace}(DI)$$
$$= \text{trace}(D)$$
$$= \lambda_1 + \lambda_2 + \lambda_3 + \ldots + \lambda_p.$$

$\therefore \sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp} = \lambda_1 + \lambda_2 + \ldots + \lambda_p$

(i.e) $\sum_{i=1}^{p} \text{var}(x_i) = \sum_{i=1}^{p} \text{var}(y_i)$

2(b)

(i) The principal components

$$Y_1 = e_1' X = 0.383 X_1 - 0.924 X_2$$

$$Y_2 = e_2' X = X_3$$

$$Y_3 = e_3' X = 0.924 X_1 + 0.383 X_2$$

(ii) Yes. X_3 is a principal component because it is uncorrelated with the other two variables (i.e X_1 & X_2)

(iii) $\text{var}(Y_1) = \text{var}(0.383 X_1 - 0.924 X_2)$

$$= (0.383)^2 \text{var}(X_1) + (-0.924)^2 \text{var}(X_2)$$

$$- 2.(0.383)(0.924) \text{cov}(x_1, x_2)$$

$$= 0.147(1) + 0.854(5) - 0.708(-2)$$

$$= 5.83 = \lambda_1 \rightarrow ①$$

$\text{var}(Y_2) = \text{var}(X_3) = \Sigma_{33} = 2 = \lambda_2 \rightarrow ②$

$$\text{var}(Y_3) = \text{var}(0.924\, X_1 + 0.383\, X_2)$$

$$= (0.924)^2\, \text{var}(X_1) + (0.383)^2\, \text{var}(X_2) + 2.(0.924)(0.383)\, \text{cov}(X_1, X_2)$$

$$= 0.853(1) + (0.1466).5 + 2 \times 0.924 \times 0.383 \times (-2)$$

$$= 0.171$$

$$= \lambda_3 \rightarrow (3)$$

$\therefore$ from (1), (2) and (3),

$$\text{var}(Y_i) = \lambda_i \quad \text{where } i = 1, 2, 3$$

Also,

$$\text{cov}(Y_1, Y_2) = \text{cov}(0.383\, X_1 - 0.924\, X_2,\ X_3)$$

$$= 0.383\, \text{cov}(X_1, X_3) - 0.924\, \text{cov}(X_2, X_3)$$

$$= 0$$

$$\text{cov}(Y_2, Y_3) = \text{cov}(X_3,\ 0.924\, X_1 + 0.383\, X_2)$$

$$= 0.924\, \text{cov}(X_3, X_1) + 0.383\, \text{cov}(X_3, X_2)$$

$$= 0$$

$$\text{cov}(Y_1, Y_3) = \text{cov}(0.383 X_1 - 0.924 X_2,\ 0.924 X_1 + 0.383 X_2)$$

$$= 0.383 \times 0.924\, \text{var}(X_1) + 0.383^2\, \text{cov}(X_1, X_2) - 0.924^2\, \text{cov}(X_1, X_2) - 0.383 \times 0.924\, \text{var}(X_2)$$

$= 0.353(1) + 0.147(-2) - 0.853(-2) - 0.353(5)$

$= 0$ //

(iv)

The ratio of first principal component to total variance $\Big\} = \dfrac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = \dfrac{5.83}{8} = 0.73$

The ratio of second principal component to total variance $\Big\} = \dfrac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \dfrac{2}{8} = 0.25$

The ratio of third principal component to total variance $\Big\} = \dfrac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} = \dfrac{0.173}{8} = 0.02$

The addition of ratio of first and second principal components $= 0.73 + 0.25 = 0.98$

Hence first two components account for 0.98 of population variance. Hence Y_1 and Y_2 could replace the three variables with little loss of information

3.

$$y^{(pr)} \sim \mathcal{N}(\mu_p, \sigma_p^2)$$

$$z^{(pr)} \sim \mathcal{N}(\nu_r, \tau_r^2)$$

$$x^{(pr)} \mid y^{(pr)}, z^{(pr)} \sim \mathcal{N}(y^{(pr)} + z^{(pr)}, \sigma^2)$$

Step 1: E-step

$$x^{(pr)} = y^{(pr)} + z^{(pr)} + \epsilon^{(pr)} \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Note: ϵ is independent of y, z

$$x^{(pr)} \sim \mathcal{N}(\mu_p + \nu_r, \sigma_p^2 + \tau_r^2 + \sigma^2)$$

For the joint distribution $(y^{(pr)}, z^{(pr)}, x^{(pr)})$

mean vector $m_{pr} = [\mu_p, \nu_p, \mu_p + \nu_r]^T$

$$\text{cov}(x, y) = \text{cov}(y + z + \epsilon, y)$$

$$= E[(y + z + \epsilon - (\mu_p + \nu_r))(y - \mu_p)]$$

$$= E[[(y - \mu_p) + (z - \nu_r) + \epsilon][y - \mu_p]]$$

$$= E[(y - \mu_p)^2] + E[(z - \nu_r)(y - \mu_p)] + E[\epsilon(y - \mu_p)]$$

$$= E[(y - \mu_p)^2] + E[z - \nu_r]\,E[y - \mu_p] + E[\epsilon]\,E[y - \mu_p]$$

since they are independent

$$= \sigma_p^2$$

Similarly $\operatorname{cov}(x,z) = \operatorname{cov}(y+z+\epsilon, z)$

$= \tau_r^2$

$\operatorname{cov}(y,z) = 0$, since they are independent

$$\Sigma_{pr} = \begin{bmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_r^2 & \tau_r^2 \\ \sigma_p^2 & \tau_r^2 & \sigma_p^2+\tau_r^2+\sigma^2 \end{bmatrix}$$

$$p(x^{(pr)}, y^{(pr)}, z^{(pr)}; \mu_p, \nu_r, \sigma_p^2, \tau_r^2) = \frac{1}{(2\pi)^{3/2} |\Sigma_{pr}|^{1/2}} \exp\left(\frac{1}{2}(a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr})\right)$$

where $a^{pr} = [y^{pr}, z^{pr}, x^{pr}]$

$Q_{pr}(y^{(pr)}, z^{(pr)}) = p(y^{(pr)}, z^{(pr)} \mid x^{(pr)})$

Let $x_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$

$x_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$

x be a random variable after combining x_1 & x_2

$x = [x_1, x_2]^T \sim \mathcal{N}([\mu_1, \mu_2]^T, \Sigma)$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$

$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

$\mu_1 = [\mu_p, \nu_r]^T$, $\Sigma_{12} = [\sigma_p^2, \tau_r^2]^T$

$\Sigma_{22}^{-1} = \dfrac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2}$, $x_2 = x^{pr}$

$\mu_2 = \mu_p + \nu_r$, $\Sigma_{11} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \tau_r^2 \end{bmatrix}$

$\Sigma_{21} = [\sigma_p^2, \tau^2]$

$$\mu_{1/2} = \begin{bmatrix} \mu_p \\ \nu_r \end{bmatrix} + \left(\frac{x^{(pr)} - \mu_p - \nu_r}{\sigma_p^2 + \tau_r^2 + \sigma^2}\right) \begin{bmatrix} \sigma_p^2 \\ \tau_r^2 \end{bmatrix}$$

$$\Sigma_{1|2} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \tau_r^2 \end{bmatrix} - \begin{bmatrix} \sigma_p^2 \\ \tau_r^2 \end{bmatrix} \frac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2} [\sigma_p^2 \quad \tau_r^2]$$

$$= \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \tau_r^2 \end{bmatrix} - \frac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2} \begin{bmatrix} \sigma_p^4 & \sigma_p^2 \tau_r^2 \\ \tau_r^2 \sigma_p^2 & \tau_r^4 \end{bmatrix}$$

$Q_{pr}(y^{(pr)}, z^{(pr)}) = p(y^{(pr)}, z^{(pr)} \mid x^{(pr)})$

$$= \frac{1}{\sqrt{2\pi} |\Sigma_{1|2}|} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} y^{(pr)} \\ z^{(pr)} \end{bmatrix} - \mu_{1|2}\right)^T \Sigma_{1|2}^{-1} \left(\begin{bmatrix} y^{(pr)} \\ z^{(pr)} \end{bmatrix} - \mu_{1|2}\right)\right)$$

Let

$$w(y^{(pr)}, z^{(pr)}) = Q_{pr}(y^{(pr)}, z^{(pr)})$$

The lower bound for log likelihood,

$$l(\mu_p, \nu_r, \sigma_p^2, \tau_r^2) = \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} Q_{pr}(y^{(pr)}, z^{(pr)}) \log \frac{p(y^{(pr)}, z^{(pr)}, x^{(pr)})}{Q_{pr}(y^{(pr)}, z^{(pr)})}$$

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} w(y^{(pr)}, z^{(pr)}) \log \frac{p(y^{(pr)}, z^{(pr)}, x^{(pr)})}{w(y^{(pr)}, z^{(pr)})}$$

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{(y,z)} w(y^{(pr)}, z^{(pr)}) \log \frac{\frac{1}{(2\pi)^{2/3} |\Sigma_{pr}|^{1/2}} \exp\left(-\frac{1}{2} (a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr})\right)}{w(y^{(pr)}, z^{(pr)})}$$

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} w(y^{(pr)}, z^{(pr)}) \left(\log \frac{1}{(2\pi)^{2/3} |\Sigma_{pr}|^{1/2}} - \frac{1}{2} (a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr}) - \log w(y^{(pr)}, z^{(pr)}) \right)$$

where

$$a^{(pr)} = [y^{(pr)}, z^{(pr)}, x^{(pr)}]^T$$

$$m_{pr} = [\mu_p, \nu_r, \mu_p + \nu_r]^T$$

$$\Sigma_{pr} = \begin{bmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_r^2 & \tau_r^2 \\ \sigma_p^2 & \tau_r^2 & \sigma_p^2 + \tau_r^2 + \sigma^2 \end{bmatrix}$$

$$|\Sigma_{pr}| = \sigma_p^2 \tau_r^2 \sigma^2$$

$$C = \begin{bmatrix} \tau_r^2(\sigma_p^2 + \sigma^2) & \sigma_p^2 \tau_r^2 & -\sigma_p^2 \tau_r^2 \\ \sigma_p^2 \tau_r^2 & \sigma_p^2(\tau_r^2 + \sigma^2) & -\sigma_p^2 \tau_r^2 \\ -\sigma_p^2 \tau_r^2 & -\sigma_p^2 \tau_r^2 & -\sigma_p^2 \tau_r^2 \end{bmatrix}$$

$$\Sigma_{pr}^{-1} = \frac{C}{|\Sigma_{pr}|} = \begin{bmatrix} \frac{1}{\sigma_p^2} + \frac{1}{\sigma^2} & \frac{1}{\sigma^2} & \frac{-1}{\sigma^2} \\ \frac{1}{\sigma^2} & \frac{1}{\sigma^2} + \frac{1}{\tau_r^2} & \frac{-1}{\sigma^2} \\ \frac{-1}{\sigma^2} & -\frac{1}{\sigma^2} & \frac{-1}{\sigma^2} \end{bmatrix}$$

$$\frac{\partial l}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} w(y^{(pr)}, z^{(pr)}) \left(\log \frac{1}{(2\pi)^{2/3} |\Sigma_{pr}|^{1/2}} - \frac{1}{2} (a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr}) - \log w(y^{(pr)}, z^{(pr)}) \right)$$

$$= \sum_{r=1}^{R} \sum_{y,z} w(y^{(ir)}, z^{(ir)}) \left(\frac{\partial m_{ir}}{\partial \mu_i}\right)^T \Sigma_{ir}^{-1} (a^{(ir)} - m_{ir})$$

$$= \sum_{r=1}^{R} \sum_{y,z} w(y^{(ir)}, z^{(ir)}) [1, 0, 1] \Sigma_{ir}^{-1} (a^{(ir)} - m_{ir})$$

$$= \sum_{r=1}^{R} \sum_{(y,z)} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, -\frac{2}{\sigma^2}\right] \left(a^{(ir)} - m_{ir}\right)$$

$$= \sum_{r=1}^{R} \sum_{y,z} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, -\frac{2}{\sigma^2}\right] a^{(ir)} -$$

$$\sum_{r=1}^{R} \sum_{y,z} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, \frac{-2}{\sigma^2}\right] m_{ir}$$

$$= \sum_{r=1}^{R} \sum_{y,z} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, \frac{-2}{\sigma^2}\right] \begin{bmatrix} y(ir) \\ z(ir) \\ x(ir) \end{bmatrix} -$$

$$\sum_{r=1}^{R} \sum_{(y,z)} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, \frac{-2}{\sigma^2}\right] \begin{bmatrix} \mu_i \\ \nu_r \\ \mu_i + \nu_r \end{bmatrix}$$

equating to zero

$$\sum_{r=1}^{R} \sum_{(y,z)} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, \frac{-2}{\sigma^2}\right] \begin{bmatrix} y(ir) \\ z(ir) \\ x(ir) \end{bmatrix} =$$

$$\sum_{r=1}^{R} \sum_{y,z} w(y(ir), z(ir)) \left[\frac{1}{\sigma_i^2}, 0, -\frac{2}{\sigma^2}\right] \begin{bmatrix} \mu_i \\ \nu_r \\ \mu_i + \nu_r \end{bmatrix}$$

$$\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{y(ir)}{\sigma_i^2} - \frac{2x(ir)}{\sigma^2}\right) = \sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{\mu_i}{\sigma_i^2} - \frac{2(\mu_i + \vartheta_r)}{\sigma^2}\right)$$

$$\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{y(ir)}{\sigma_i^2} - \frac{2(x(ir) - \vartheta_r)}{\sigma^2}\right) = \sum_{r=1}^{R}\sum_{(y,z)} w(y(ir), z(ir)) \left(\frac{1}{\sigma_i^2} - \frac{2}{\sigma^2}\right)\mu_i$$

$$\Rightarrow \boxed{\mu_i = \frac{\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{y(ir)}{\sigma_i^2} - \frac{2(x(ir) - \vartheta_r)}{\sigma^2}\right)}{\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{1}{\sigma_i^2} - \frac{2}{\sigma^2}\right)}}$$

$$\frac{\partial l}{\partial \sigma_i^2} = \frac{\partial}{\partial \sigma_i^2}\sum_{p=1}^{P}\sum_{r=1}^{R}\sum_{y,z} w(y(pr), z(pr))$$

$$\left(\log \frac{1}{(2\pi)^{2/3}|\Sigma_{pr}|^{1/2}} - \frac{1}{2}(a(pr) - m_{pr})\Sigma_{pr}^{-1}(a(pr) - m_{pr}) - \log w(y(pr), z(pr))\right)$$

$$= \frac{\partial}{\partial \sigma_i^2}\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{-1}{2}\log|\Sigma_{ir}| - \frac{1}{2}(a(ir) - m_{ir})^T\Sigma_{ir}^{-1}(a(ir) - m_{ir})\right)$$

$$= \frac{-1}{2}\sum_{r=1}^{R}\sum_{y,z} w(y(ir), z(ir))\left(\frac{1}{|\Sigma_{ir}|}\cdot\frac{\partial|\Sigma_{ir}|}{\partial\sigma_i^2} - \frac{1}{2}(a(ir) - m_{ir})^T\Sigma_{ir}^{-1}(a(ir) - m_{ir})\right)$$

$$= \frac{-1}{2} \sum_{r=1}^{R} \sum_{y,z} w_{(y^{(ir)}, z^{(ir)})} \left(\frac{1}{|\Sigma_{ir}|} \frac{\partial |\Sigma_{ir}|}{\partial \sigma_i^2} + (a^{(ir)} - m_{ir})^T \frac{\partial \Sigma_{ir}^{-1}}{\partial \sigma_i^2} (a^{(ir)} - m_{ir}) \right)$$

$$= \frac{-1}{2} \sum_{r=1}^{R} \sum_{y,z} w_{(y^{(ir)}, z^{(ir)})} \left(\frac{1}{\sigma_i^2} + \begin{bmatrix} y^{(ir)} - \mu_i \\ z^{(ir)} - \nu_r \\ x^{(ir)} - \mu_i - \nu_r \end{bmatrix}^T \begin{bmatrix} -\frac{1}{\sigma_i^4} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y^{(ir)} - \mu_i \\ z^{(ir)} - \nu_r \\ x^{(ir)} - \mu_i - \nu_r \end{bmatrix} \right)$$

$$= \frac{-1}{2} \sum_{r=1}^{R} \sum_{y,z} w_{(y^{(ir)}, z^{(ir)})} \left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^4} (y^{(ir)} - \mu_i)^2 \right)$$

substituting to zero, we get

$$\boxed{\sigma_i^2 = \frac{\sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{(ir)}, z^{(ir)})} (y^{(ir)} - \mu_i)^2}{\sum_{r=1}^{R} \sum_{y,z} w_{(y^{(ir)}, z^{(ir)})}}}$$

$$\frac{\partial l}{\partial \nu_j} = \frac{\partial}{\partial \nu_j} \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} w_{(y^{(pr)}, z^{(pr)})} \left(\log \frac{1}{(2\pi)^{3/2} |\Sigma_{pr}|^{1/2}} - \frac{1}{2} (a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr}) - \log w_{(y^{(pr)}, z^{(pr)})} \right)$$

$$= \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left(\frac{\partial m_{pj}}{\partial \nu_j}\right) \Sigma_{pj}^{-1} (a^{(pj)} - m_{pj})$$

$$= \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) [0,1,1] \Sigma_{pj}^{-1} (a^{(pj)} - m_{pj})$$

$$= \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, -\frac{2}{\sigma^2}\right] (a^{(pj)} - m_{pj})$$

$$= \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] a^{(pj)} - \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] m_{pj}$$

$$= \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] \begin{bmatrix} y^{(pj)} \\ z^{(pj)} \\ x^{(pj)} \end{bmatrix} - \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] \begin{bmatrix} \mu_p \\ \nu_j \\ \mu_p + \nu_j \end{bmatrix}$$

Equating to zero,

$$\sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] \begin{bmatrix} y^{(pj)} \\ z^{(pj)} \\ x^{(pj)} \end{bmatrix} = \sum_{p=1}^{P} \sum_{y,z} \left[0, \frac{1}{\tau_j^2}, \frac{-2}{\sigma^2}\right] \begin{bmatrix} \mu_p \\ \nu_j \\ \mu_p + \nu_j \end{bmatrix}$$

$, z^{(pj)}))$

$$\sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{z^{(pj)}}{\tau_j^2} - \frac{2x^{(pj)}}{\sigma^2} \right) =$$

$$\downarrow$$

$$\sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{\vartheta_j}{\tau_j^2} - \frac{2(\mu_p + \vartheta_j)}{\sigma^2} \right)$$

$$\sum_{p=1}^{P} \sum_{(y,z)} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{z^{(pj)}}{\tau_j^2} - \frac{2(x^{(pj)} - \mu_p)}{\sigma^2} \right) =$$

$$\downarrow$$

$$\sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{1}{\tau_j^2} - \frac{2}{\sigma^2} \right) \vartheta_j$$

$$\boxed{\vartheta_j = \frac{\displaystyle\sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{z^{(pj)}}{\tau_j^2} - \frac{2(x^{(pj)} - \mu_p)}{\sigma^2} \right)}{\displaystyle\sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{1}{\tau_j^2} - \frac{2}{\sigma^2} \right)}}$$

$$\frac{\partial l}{\partial \tau_j^2} = \frac{\partial}{\partial \tau_j^2} \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{y,z} w_{(y^{(pr)}, z^{(pr)})}$$

$$\left(\log \frac{1}{(2\pi)^{2/3} |\Sigma_{pr}|^{1/2}} - \frac{1}{2}(a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{(pr)} - m_{pr}) - \log w(y^{(pr)}, z^{(pr)}) \right)$$

$$= \frac{\partial}{\partial \tau_j^2} \sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(-\frac{1}{2} \log |\Sigma_{pj}| - \frac{1}{2}(a^{(pj)} - m_{pj})^T \Sigma_{pj}^{-1} (a^{(pj)} - m_{pj}) \right)$$

$$= -\frac{1}{2} \sum_{p=1}^{P} \sum_{y,z} w_{(y^{(pj)}, z^{(pj)})} \left(\frac{1}{|\Sigma_{pj}|} \frac{\partial |\Sigma_{pj}|}{\partial \tau_j^2} + (a^{(pj)} - m_{pj})^T \frac{\partial \Sigma_{pj}^{-1}}{\partial \tau_j^2} (a^{(pj)} - m_{pj}) \right)$$

$$= -\frac{1}{2} \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left(\frac{1}{\tau_j^2} + \begin{bmatrix} y^{(pj)} - \mu_p \\ z^{(pj)} - \nu_j \\ x^{(pj)} - \mu_p - \nu_j \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{\tau_j^4} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y^{(pj)} - \mu_p \\ z^{(pj)} - \nu_j \\ x^{(pj)} - \mu_p - \nu_j \end{bmatrix} \right)$$

$$= -\frac{1}{2} \sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)}) \left(\frac{1}{\tau_j^2} - \frac{1}{\tau_j^4} (z^{(pj)} - \nu_j)^2 \right)$$

$$\boxed{\tau_j^2 = \frac{\sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)} - z^{(pj)}) \left(z^{(pj)} - \nu_j\right)^2}{\sum_{p=1}^{P} \sum_{y,z} w(y^{(pj)}, z^{(pj)})}}$$