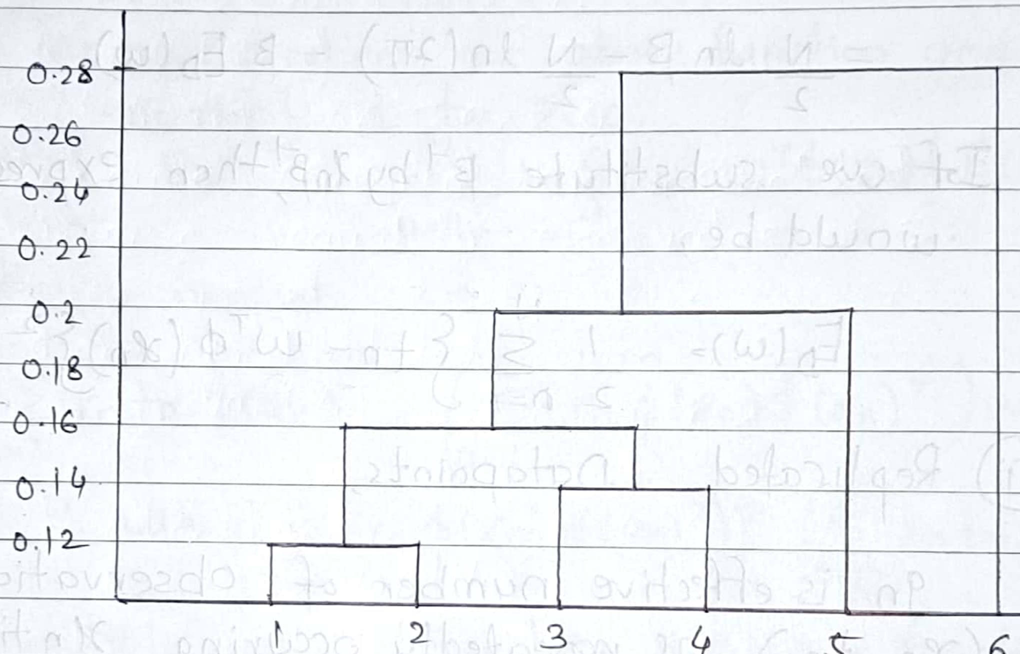
Assignment-5

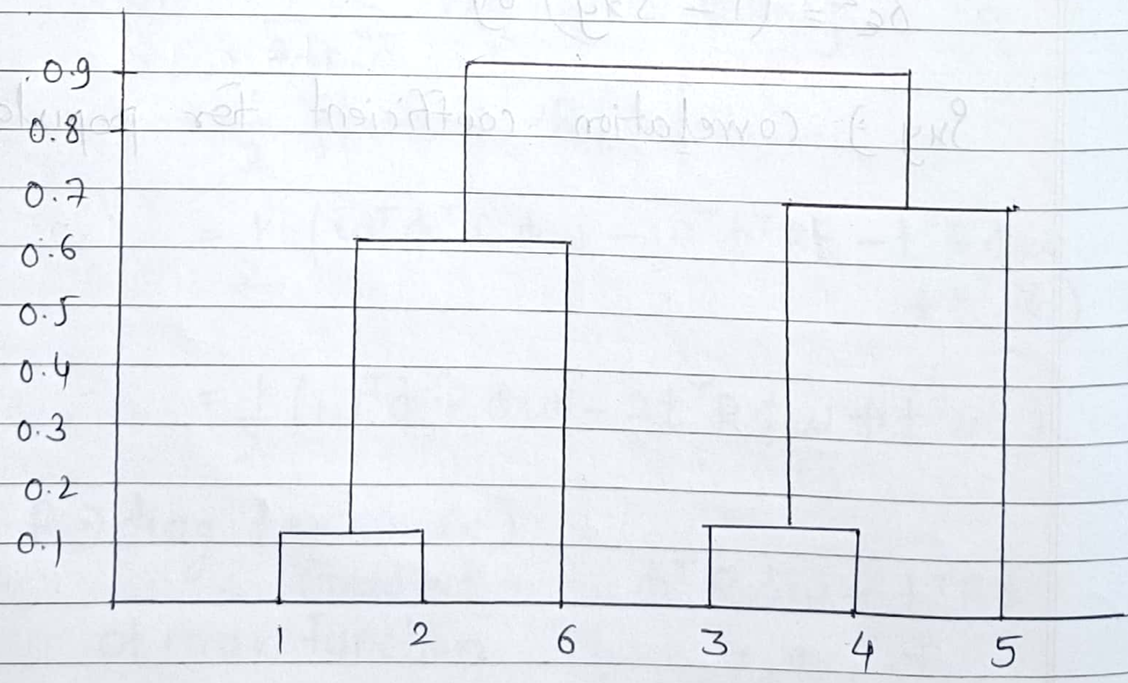
Foundations of Machine Learning

Q1) Hierarchical Clustering

a) Final result of hierarchical clustering with Single link



b) Final result with complete link.



c) 1. Complete link clustering differs from single link clustering is where AB and F are grouped together by distance (AB,F) = distance (B,F) = 0.61. We should want dist (AB, CD) = dist (A,D) to be smaller than 0.61, such as 0.53.

2. Second step is, we want dist (ABCD, F) = dist (C,F) = 0.93 to be the smallest so that ABCD and F grouped together. We can set this value to 0.63. After changing these values, both dendograms become identical.

Q.2

a) $X' = [X_1, X_2 \ldots X_p]$ cov. matrix $\Rightarrow \Sigma$

eigen val - eigen vector $\Rightarrow$ $(\lambda_1, e_1), (\lambda_2, e_2) \ldots (\lambda_p, e_p)$

$Y_1 = e_1' X, \; Y_2 = e_2' X, \ldots Y_p = e_p' X$. Principal components.

Prove:

$$\sigma_{11} + \sigma_{22} + \ldots \sigma_{pp} = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 \ldots \lambda_p = \sum_{i=1}^{p} Var(Y_i)$$

$$Y_i = e_i^t X = e_{i1} X_1 + e_{i2} X_2 + \ldots e_{ip} X_p \quad i = 1, 2, \ldots p.$$

$$Var(Y_i) = e_i^t \Sigma e_i = \lambda_i \qquad i = 1, 2, \ldots p$$

$$Cov(Y_i, Y_k) = e_i^t \Sigma e_k \quad i \neq k$$

From matrix algebra.

$$\max_{a \neq 0} \frac{a^t \Sigma a}{a^t a} = \lambda_1$$

When $a = e_1$, $e_1^t e_1 = 1$ Hence,

$$\max_{a \neq 0} \frac{a^t \Sigma a}{a^t a} = \lambda_1 = \frac{e_1^t \Sigma e_1}{e_1^t e_1} = e^t \Sigma e_1 = Var(Y_1)$$

Simi, $\max\limits_{a \perp e_1, e_2 \ldots e_k} \frac{a^t \Sigma a}{a^t a} = \lambda_{k+1} \quad k = 1, 2, \ldots p-1$

$$\frac{e_{k+1}^t \Sigma e_{k+1}}{e_{k+1}^t e_{k+1}} = e_{k+1}^t \Sigma e_{k+1} = Var(Y_{k+1}).$$

But $e_{k+1}^t \Sigma e_{k+1} = \lambda_{k+1} e_{k+1}^t e_{k+1} = \lambda_{k+1}$

$\therefore Var(Y_{k+1}) = \lambda_{k+1} \quad -\mathrm{I}$

$Cov(Y_i, Y_k) = e_i^t \Sigma e_k = 0$

$\therefore$ Principle components are uncorrelated and have variance equal to eigenvalue of Σ.

On one hand,

$$\sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} \sigma_{ii} = tr\,\Sigma = \sum_{i=1}^{p} \lambda_i \quad -\mathrm{II}$$

On the other hand,

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} \lambda_i \quad -\mathrm{III}$$

$$\therefore \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i) \ldots \text{ from } \mathrm{II} \,\&\, \mathrm{III}$$

is proved

classmate

Date

Page

b) $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $\lambda_1 = 5.83$, $e_1' = [0.383, -0.924, 0]$
$\lambda_2 = 2.00$, $e_2' = [0, 0, 1]$
$\lambda_3 = 0.17$, $e_3' = [0.924, 0.383, 0]$

i) Principle 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) X_3 is one of the principle component, since it is uncorrelated with the other two variables.

iii) $Var(Y_i) = \lambda_i$, $Cov(Y_i, Y_k) = 0$, $i \neq k$.

$Var(Y_i) = e_i' \Sigma e_i'$, $Cov(Y_i Y_k) = e_i' \Sigma e_k' = 0$

$Var(Y_i) = Var(0.383 X_1 - 0.924 X_2)$

$= (0.383)^2 \; Var(X_1) + (-0.924)^2 \; Var(X_2) + 2(0.383)(-0.924) \; Cov(X_1 X_2)$

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

$\therefore Var(Y_i) = 5.83 = \lambda_1$ ... proved

$Cov(Y_1 Y_2) = Cov(0.383 X_1 - 0.924 X_2, X_3)$

$= 0.383 \; Cov(X_1 X_3) - 0.924 \; Cov(X_2 X_3)$

$= 0.383(0) - 0.924(0)$

$\therefore Cov(Y_1 Y_2) = 0$ ... proved classmate

Date
Page

iv) $\sigma_{11}+\sigma_{22}+\sigma_{33} = 1+5+2 = \lambda_1+\lambda_2+\lambda_3 = 5.83+2+0.17$

$\therefore$ Total variance $= \sigma_{11}+\sigma_{22}+\dots\sigma_{pp}$

$= \lambda_1+\lambda_2+\dots+\lambda_p$

Fraction of total variance accounted for by first principle component is $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = \frac{5.83}{8} = 0.73$

First two principle components $\Rightarrow \frac{5.83+2}{8} = 0.98$ of population variance

Components Y_1 & Y_2 could restore initial 3 variables with little loss of information

$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$ $i, k = 1, 2, \dots p$

$\therefore \rho_{Y_1, X_1} = 0.925$, $\rho_{Y_1, X_2} = -0.998$

Here variable X_2 with coefficient -0.924 obtains most wt. in component Y_1. It has largest correlation with Y_1 (abs. value). The correlation of X_1 and Y_1, 0.925 is roughly as large as that of X_2; the variables are almost equally important to the first principal component.

Lastly,

$\rho_{Y_2 X_1} = \rho_{Y_2 X_2} = 0$ and $\rho_{Y_2 X_3} = \frac{\sqrt{2}}{\sqrt{2}} = 1$

The third component is irrelevant and because of this, the remaining correlations can be neglected.

Q.3) EM Application

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

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

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

$y^{(pr)}, z^{(pr)} \ni$ independent

a) E-step :

i) $x^{(pr)} = y^{(pr)} + z^{(pr)} + \varepsilon^{(pr)}$, $\varepsilon \sim N(0, \sigma^2)$

$x^{(pr)} \ni$ a normal distribution that is the sum of multiple independent normal distributions.

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

For the joint distribution, $p(y^{(pr)}, z^{(pr)}, x^{(pr)})$, its mean vector and covariance matrix are

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

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

$$\begin{aligned} \operatorname{cov}(A, A+B) &= E[(A - E[A])(A + B - E[A+B])] \\ &= E[(A - E[A])(A - E[A] + B - E[B])] \\ &= E[(A - E[A])(A - E[A]) + (A - E[A])(B - E[B])] \\ &= E[(A - E[A])(A - E[A])] \\ &= \sigma_A^2 \end{aligned}$$

Therefore, a trivariate normal distribution,

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

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

ii) $x_1 \sim N(\mu_1, \Sigma_{11})$, $x_2 \sim N(\mu_2, \Sigma_{22})$

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

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

x_1, x_2, μ_1, μ_2 and σ can not only be scalars, but also vectors / submatrices

Then for conditional random variable,

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$
$$\Sigma_{1|2} = \Sigma_{11} + \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

$$\mu_1 = [\mu_p, V_r]^T$$

$$\Sigma_{12} = [\sigma_p^2, \tau^2]^T$$

$$\Sigma_{22}^{-1} = \frac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2}$$

$$x_2 = x^{(pr)}$$
$$\mu_2 = \mu_p + V_r$$

$$\Sigma_{11} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \tau_r^2 \end{bmatrix}, \quad \Sigma_{21} = [\sigma_p^2, \tau^2]$$

So,

$$\mu_{1|2} = \begin{bmatrix} \mu_p \\ V_r \end{bmatrix} + \frac{x^{(pr)} - \mu_p - V_r}{\sigma_p^2 + \tau_r^2 + \sigma^2} \begin{bmatrix} \sigma_p^2 \\ \tau^2 \end{bmatrix}$$

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

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

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

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

Q_{pr} follows a bivariate normal distribution.

b) <u>M Step:</u>

$$W\left(y^{(pr)}, z^{(pr)}\right) = Q_{pr}\left(y^{(pr)}, z^{(pr)}\right)$$

Lower bound for log likelihood be:

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

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

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

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

Listed all parts of above equation,

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

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

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

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

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

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

To update the true value of the i^{th} paper, μ_i, maximize the likelihood w.r.t μ_i

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

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

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

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

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

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

Equating above equation to zero we get,

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

Updating the variance of μ_i i.e. σ_i

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

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

Equating to zero we get,

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

Updating the bias of j^{th} reviewer,

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

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

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

Equating to zero,

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

Updating the variance of v_j,

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

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

Equating to zero,

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

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

Summarizing the Updates,

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

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

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

iv) $$T_j^2 = \frac{\sum_{p=1}^{P} \sum_{(y,z)} w_{(y(pj)),\, z(pj)} \left(z^{(pj)} - v_j \right)^2}{\sum_{p=1}^{P} \sum_{(y,z)} w(y(pj), z(pj))}$$

Derivations of μ_i, σ_i^2, v_j, T_j are analogous, given one derivation, the others are apparent.