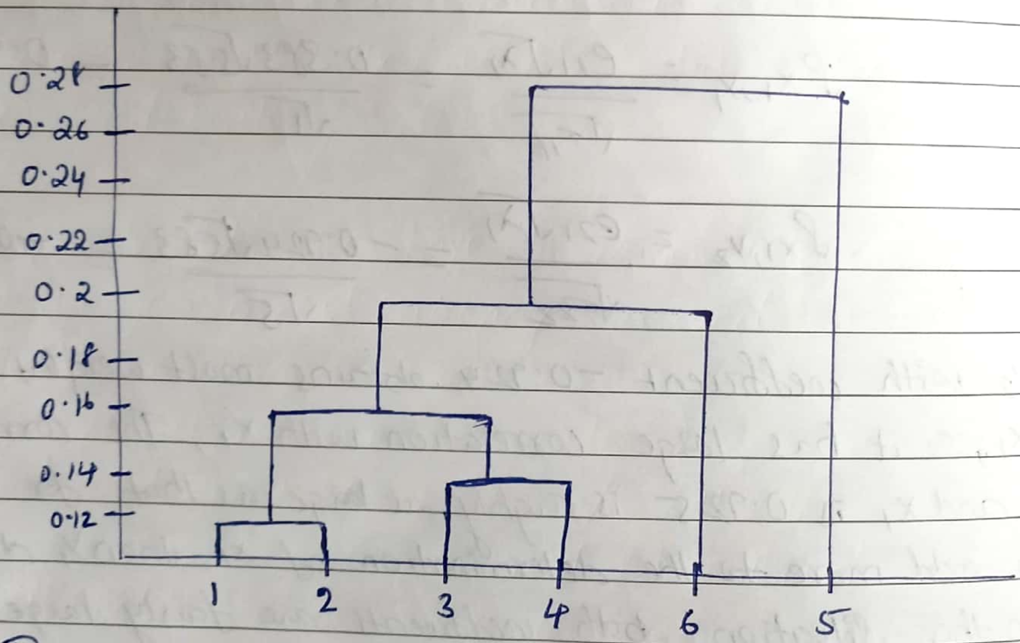
FoML - Assignment-5

ES19BTECH11003
Hemanth K

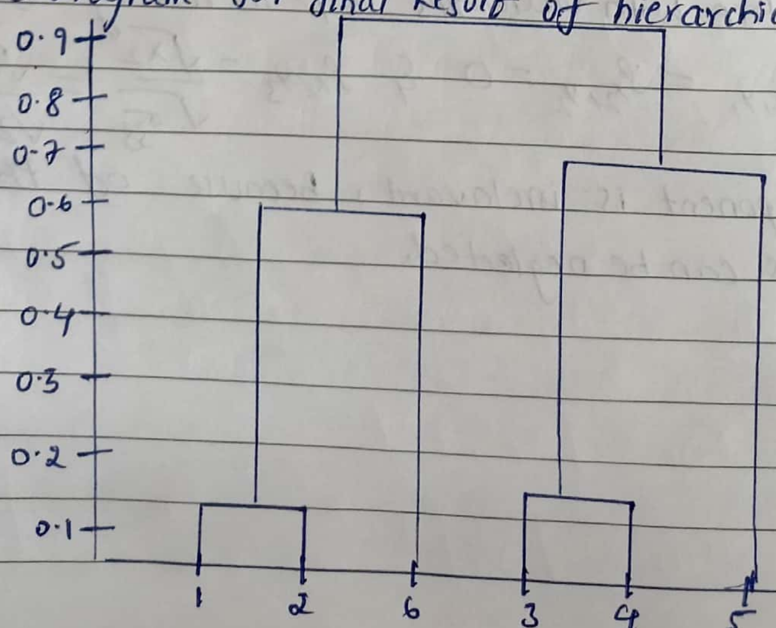
1) Given, distance matrix for 6 data points

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0.12	0				
x_3	0.51	0.25	0			
x_4	0.84	0.16	0.14	0		
x_5	0.28	0.77	0.70	0.45	0	
x_6	0.34	0.61	0.93	0.20	0.67	0

a) Dendrogram for the final Result of hierarchical clustering with single link.



b) Dendrogram for final Result of hierarchical clustering with complete Link



c) The main difference where the complete link clustering and single link differs is where AB and F are grouped together by dist(AB, F) = dist(B, F) = 0.61, so we would want dist(AB, CD) = dist(A, D) to be smaller than this value such as 0.56, then we want dist(ABCD, F) = dist(C, F) = 0.93 to be the smallest so that ABCD & F are grouped together. we set this value to 0.63. After these changes both dendrograms become identical

2)

a) Given, $x' = [x_1, x_2, \dots x_p]$ have covariance matrix Σ,
eigenvalue-eigen vector pairs $(\lambda_1, e_1), (\lambda_2, e_2) \dots (\lambda_p, e_p)$
where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$

to prove:- $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + \dots \lambda_p = \sum_{i=1}^{p} Var(Y_i)$

proof:-

$Y_1 = e_1^t x$, $Y_2 = e_2^t x$, $\dots$ $Y_p = e_p^t x$, be the principal components

$$\sigma_{11} + \sigma_{22} + \dots \sigma_{pp} = \sum_{i=1}^{p} Var(Y_i)$$

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

also $\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} \lambda_i$

The proportion of total variance due to (explained by) the k^{th} principal component is

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots \lambda_p}, \quad k = 1, 2, \dots p$$

b) Given, random Variables X_1, X_2, X_3 have covariance matrix

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\lambda_1 = 5.83, \quad e_1' = [0.383, -0.924, 0]$$
$$\lambda_2 = 2.00, \quad e_2' = [0, 0, 1]$$
$$\lambda_3 = 0.17, \quad e_3' = [0.924, 0.383, 0]$$

to find:-

principal components Y_1, Y_2, Y_3

$$X_1 = e_1'Y = 0.383Y_1 - 0.924Y_2$$
$$X_2 = Y_3$$
$$X_3 = e_3'Y = 0.924Y_1 - 0.383Y_2$$

Y_3 is one of the principle component, as it is uncorrelated with other two variables

$$Var(X_i) = e_i' \Sigma e_i' = \lambda_i, \quad i = 1, 2, \ldots p \quad — ①$$
$$Cov(X_i X_k) = e_i' \Sigma e_k' = 0 \quad i \neq k \quad — ②$$

① − ②

$$Var(X_1) = Var(0.383Y_1 - 0.924Y_2)$$
$$= (0.383)^2 Var(Y_1) + (-0.924)^2 Var(Y_2)$$
$$-2(0.383)(0.924) Cov(Y_1, Y_2)$$
$$Var(X_1) = 0.147 Var(Y_1) + 0.854(Var(Y_2)) - 0.708(-2)$$
$$= 0.147(1) + 0.854(5) - 0.708(-2)$$
$$Var(X_1) = 5.83 = \lambda_1$$
$$Cov(X_1, X_2) = Cov(0.383Y_1 - 0.924Y_2, Y_3) = 0.383 Cov(Y_1, Y_3)$$
$$- 0.924 Cov(Y_2, Y_3)$$
$$= 0.383(0) - 0.924(0) = 0$$

Hence $\sigma_{11} + \sigma_{22} + \sigma_{33} = 1+5+2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2.00 + 0.17$

total Variance $= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$

$$= \lambda_1 + \lambda_2 + \lambda_3 \cdots \lambda_p$$

fraction of total variance because of the first principal component

$$= \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = 5.83/8 = 0.73$$

the first 2 components $= \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \frac{5.83+2}{8} = 0.98$

In this situation, components x_1, x_2 could restore the initial 3 variables with little loss of information

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

$$\rho_{x_1,y_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{0.383\sqrt{5.83}}{\sqrt{1}} = 0.925$$

$$\rho_{x_1,y_2} = \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-0.924\sqrt{5.83}}{\sqrt{5}} = -0.998$$

Y_2 with coefficient -0.924 obtains most weight in component X_1, so it has large correlation with x_1. The correlation of Y_1 and X_1 is 0.925 is roughly as large as that for Y_2.
Y_2 add more to the determination of x_1 than Y_1 does. But in this situation, both coefficients are fairly large & of opposite signs, therefore both variables help in reading of x_1,

$$\rho_{x_2,y_1} = \rho_{x_2,y_2} = 0 \text{ \& } \rho_{x_2 y_3} = \frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} = 1$$

$\therefore$ 3rd component is irrelevant, because of this remaining correlations can be neglected.

Scanned with CamScanner

3) Given, we assume reviewer's scores are generated by a random process $y^{(pr)} \sim \mathcal{N}(\mu_p, \sigma_p^2)$

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

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

y^{pr} & z^{pr} are independent

y^{pr}'s and z^{pr}'s are all latent random variables.

~~a) to find joint distribution~~ ..

a) Find associated mean vector & co-variance matrix for the joint distribution $p(y^{pr}, z^{pr}, x^{pr})$ in terms of parameters $\mu_p, \sigma_p^2, \nu_r, \tau_r^2$ & σ^2

Proof:-

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

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

for joint distribution $p(y^{pr}, z^{pr}, x^{pr})$ & its mean vector μ_{pr} covariance matrix σ_{pr}

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

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

$Cov(y^{pr}, x^{pr})$ & $cov(z^{pr}, x^{pr})$ are derived based on following. If normally distributed random variables A & B are independent, then $cov(A, A+B) = cov(A, A) = \sigma_A$

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

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

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

(ii) <u>to prove</u>:- to derive expression for $Q_{pr}(y^{pr}, z^{pr}) = p(y^{pr}, z^{pr} | x^{pr})$

<u>proof</u>:- consider x_1, x_2 be 2 multivariate normal random variables

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

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

Let x be new multivariate random variable after stacking x_1 & x_2

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

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

Notes:- $x_1, x_2, \mu_1, \mu_2, \sigma_{ij}$ cannot only be scalars, but also vectors/submatrices

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

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

we derive expression for $Q_{pr}(y_{pr}, z_{pr}) = p(y^{pr}, z^{pr} | x^{pr})$

$$\mu_1 = [\mu_p, v_r]^T \qquad \Sigma_{12} = [\sigma_p^2, \tau^2]^T$$

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

$$x_2 = x^{(pr)}$$

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

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

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

$$\Sigma_{1/2} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & T_r^2 \end{bmatrix} - \begin{bmatrix} \sigma_p^2 \\ T^2 \end{bmatrix} \frac{1}{\sigma_p^2 + T_r^2 + \sigma^2} [\sigma_p^2, T^2]$$

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

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

$$Q_{pr}(y^{pr}, z_p^{pr}) = \frac{1}{2\pi |\Sigma_{1/2}|} \exp\left(-\frac{1}{2}\left(\left[\begin{matrix} y^{pr} \\ z^{pr} \end{matrix}\right] - \mu_{1/2}\right)^T \Sigma_{1/2}^{-1} \left(\left[\begin{matrix} y^{pr} \\ z^{pr} \end{matrix}\right] - \mu_{1/2}\right)\right)$$

b) to prove :- to derive M Steps updates for parameters $(\mu_p, \sigma_p^2, v_r, T_r^2)$

Lower bound $\quad l(\mu_p, v_r, \sigma_p^2, T_r^2) = \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{(y,z)} Q_{pr}(y^{pr}, z^{pr}) \log \frac{P(y^{pr}, z^{pr}, x^{pr})}{Q_{pr}(y^{pr}, z^{pr})}$

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

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

We know $\quad a^{pr} = [y^{pr}, z^{pr}, x^{pr}]^T$

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

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

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

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

C is co-factor matrix

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

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

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

We set it to 0 we get

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

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

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

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

Set it to 0, we get

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

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

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

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

Set it to 0 we get

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

Now,

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

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

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

Set it to 0, we get

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

finally, the expressions are

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

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

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

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