Assignment-5

(1) Distance Matrix is given.

	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) single link. min nearest.

0.12 is minimum distance present b/w. x_1 & x_2

Updated matrix.

$\min[(x_1,x_3),(x_2,x_3)] = 0.25$

$\min[(x_1,x_4),(x_2,x_4)] = 0.16$

$\min[(x_1,x_5),(x_2,x_5)] = 0.28$

$\min[(x_1,x_6),(x_2,x_6)] = 0.34.$

	x_1 and x_2	x_3	x_4	x_5	x_6
x_1 & x_2	0				
x_3	0.25	0			
x_4	0.16	0.14	0		
x_5	0.28	0.70	0.45	0	
x_6	0.34.	0.93.	0.20	0.67	0.

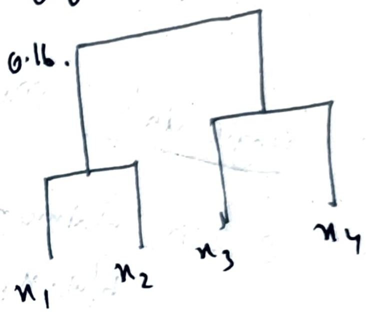
Updated Matrix

$\min[(x_1 \& x_2, x_3), (x_1 \& x_2, x_4)] : 0.16$

$\min[(x_3, x_5), (x_4, x_5)] : 0.45$

$\min[(x_5, x_6), (x_4, x_6)] : 0.20$

	$x_1 \& x_2$	$x_3 \& x_4$	x_5	x_6
$x_1 \& x_2$	0			
$x_3 \& x_4$	0.16	0		
x_5	0.28	0.45	0	
x_6	0.34	0.20	0.67	0

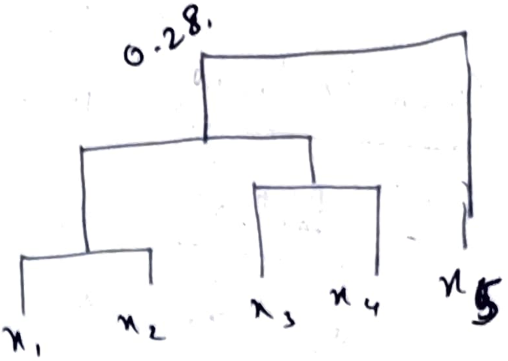


Updated Matrix

$\min[(x_1 \& x_2, x_5) (x_3 \& x_4, x_5)] = 0.28$

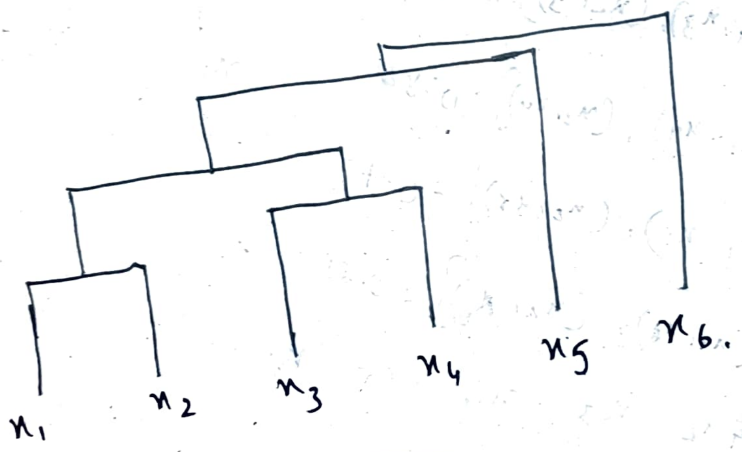
$\min[(x_1 \& x_2, x_6) (x_3 \& x_4, x_6)] : 0.20$

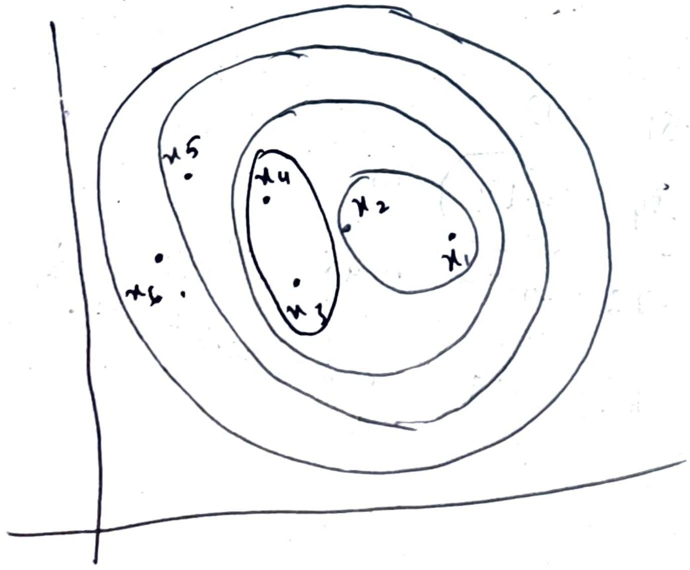
	(x_1, x_2, x_3, x_4)	x_5	x_6
(x_1, x_2, x_3, x_4)	0		
x_5	0.28	0	
x_6	0.28	0.67	0



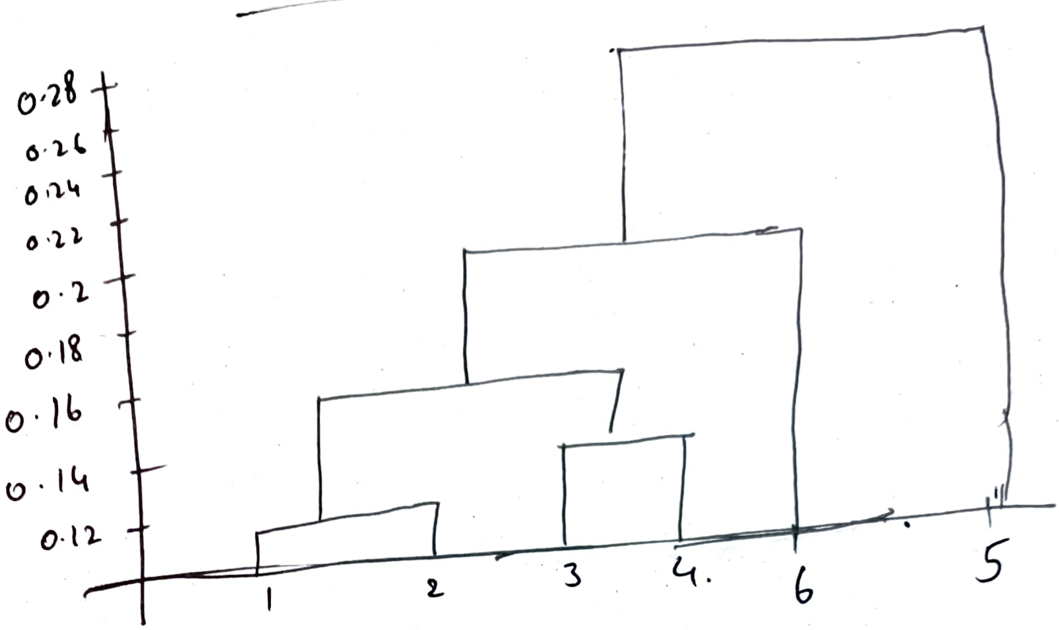
$[(x_1, x_2, x_3, x_4) \,\&\, x_6], [x_5, x_6] = \underline{\underline{0.28}}.$

	$(x_1, x_2, x_3, x_4, x_5)$	x_6
$(x_1, x_2, x_3, x_4, x_5)$	0	
x_6	0.28	0



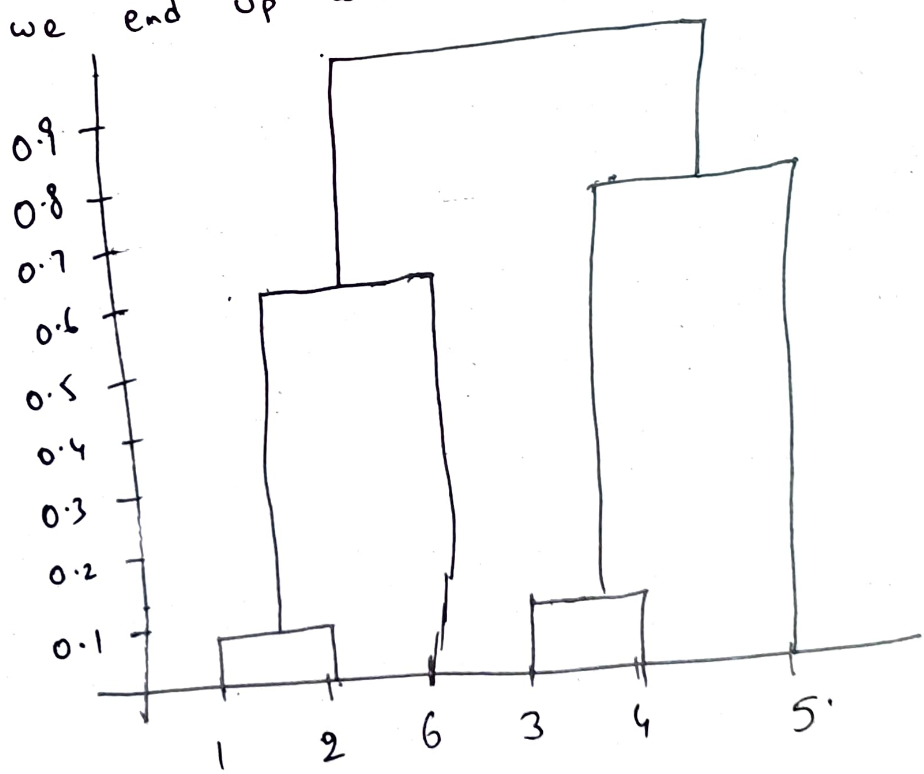


Dendogram of Single link.



Dendogram of Complete link.

After following same procedure followed for single link we end up with.



c) After changing two values dendogram should be same

for (a) Single link.

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0.13	0				
x_3	0.51	0.46	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

Value changed from 0.12 to 0.13

Value changed from 0.25 to 0.46

first cluster will be x_1 & x_2 even value is changed from 0.12 to 0.13.

Anyway 0.46 will get eliminated in further steps keeping dendogram unchanged.

for (b) Max Link.

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0.12	0				
x_3	0.8	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	0.75	0.93	0.20	0.67	0

Value chaged from 0.51 to 0.8

Value chaged from 0.61 to 0.75

After (x_3 & x_4) got clustered, in updated matrix 0.84 comes without affecting Dendogram.

Similarly 0.75 has nothing to do with existing Dendogram.

2) a) $X' = [X_1, X_2 - - - X_p]$.

Covariance Matrix Σ

Eigen value vector pairs $(\lambda_1, e_1), (\lambda_2, e_2) ... (\lambda_p, e_p)$

$\lambda_1 \geq \lambda_2 \geq ... \lambda_p \geq 0.$

$Y_1 = e_1'x \quad Y_2 = e_2'x ..... Y_p = e_p'x$ are Principal Components.

$\Rightarrow$ We know

$\sigma_{11} + \sigma_{22} + - - - - \sigma_{pp} = tr(\Sigma)$

$A = \Sigma$

$\Sigma = P \Lambda P'$

Λ is diagonal matrix of eigen values and

$P = [e_1, e_2, - - - . e_p]$

$P'P = PP' = I$

Therefore

$tr(\Sigma) = tr(P A P') = tr(P'AP) = tr(AP'P)$

$= tr(A) = \lambda_1 + \lambda_2 + - - - + \lambda_p$

$\therefore \sum_{i=1} Var(X_i) = tr(\Sigma) = tr(A) = \sum_{i=1}^{p} Var(Y_i)$

Therefore

$\sigma_{11} + \sigma_{22} + . . . + \sigma_{pp}$

$= \sum_{i=1}^{n} Var(X_i) = \lambda_1 + \lambda_2 + ... + \lambda_p$

$= \sum_{i=1}^{n} Var(Y_i)$

2b)

i) ~~Y_1~~

Co-variance matrix = $\begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Need to find Eigen Values & Eigen Vectors. (Used python program to find Eigen values & Eigen vectors).

Eigen values = $[0.17 \quad 5.82 \quad 2]$

Eigen vectors = $\begin{bmatrix} -0.923 & 0.382 & 0 \\ -0.382 & -0.923 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Y_1, Y_2, Y_3 are Principal Components

$Y_1 = \begin{bmatrix} 0.382 \\ -0.923 \\ 0 \end{bmatrix}$ Corresponding to Eigen value 5.82

$Y_1 = 0.382X_1 - 0.923X_2$

$Y_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ corresponding to Eigen value 2

$Y_2 = X_3$

$Y_3 = \begin{bmatrix} -0.923 \\ -0.382 \\ 0 \end{bmatrix}$ corresponding to Eigen value 0.17

$Y_3 = -0.923X_1 - 0.382X_2$

2b

ii) The Variable X_3 is a Principal component because it is uncorrelated with the other two variables.

iii)

$$Var(Y_1) = Var(0.382X_1 - 0.923X_2)$$

$$= (0.382)^2 Var(X_1) + (0.923)^2 Var(X_2)$$

$$+ 2(0.382)(-0.923) Cov(X_1, X_2)$$

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

$$= 5.83 = \lambda_1$$

$$\underline{Var(Y_1) = 5.83 = \lambda_1}$$

$$\underline{Var(Y_2) = Var(X_3) = 2 = \lambda_2}$$

$$Var(Y_3) = Var(0.923X_1 + 0.382X_2)$$

$$= (0.923)^2 Var(X_1) + (0.382)^2 Var(X_2)$$

$$+ 2(0.923)(0.382) Cov(X_1, X_2)$$

$$= 0.853776 + 0.733 + (-1.415568)$$

$$= 0.17 = \lambda_3$$

$$\underline{Var(Y_3) = 0.17 = \lambda_3}$$

To proove

$Cov(Y_i, Y_k) = 0$ for $i \neq k$

$$Cov(Y_1, Y_2) = Cov(0.382X_1 - 0.923X_2, X_3)$$

$$= 0.382\, Cov(X_1, X_3) - 0.923\, Cov(X_2, X_3)$$

$$= 0.382(0) - 0.923(0) = 0$$

$$\boxed{Cov(Y_1, Y_2) = 0}$$

||ly

$$Cov(Y_2, Y_3) = Cov(X_3, 0.923X_1 + 0.382X_2)$$

$$= 0.923\, Cov\underbrace{\cancel{(X_3, X_1)}}_{\to 0} + 0.383\, Cov\underbrace{\cancel{(X_3, X_2)}}_{\to 0}$$

$Cov(Y_2, Y_3) = 0$

$$Cov(Y_1, Y_3) = Cov[(0.382X_1 - 0.923X_2), (0.923X_1 + 0.382X_2)]$$

$$= (0.382)(0.923)\, Cov(X_1, X_1) + 0.382^2\, Cov(X_1, X_2) - 0.923^2\, Cov(X_2, X_1) - 0.923 \times 0.382\, Cov(X_2, X_2)$$

$$= 0.353892(1) + 0.145924(-2) - 0.851929(-2) - 0.352586(5) = 0.$$

$\therefore Cov(Y_1, Y_3) = 0.$

iv)

It is also readily apparent that

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

$$= 5.83+2.00+0.17$$

The proportion of total variance accounted for by the first principal components is $(\lambda_1)/(\lambda_1+\lambda_2+\lambda_3) = \frac{5.83}{8} = 0.73.$

Further the first two components account for a proportion $(5.83+2)/8 = 0.98$ of the population variance.

In this case, the components Y_1 & Y_2 could replace the original three variables with littles loss of information.

$$\rho_{Y_1, X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{0.382\sqrt{5.83}}{\sqrt{1}} = 0.925$$

$$\rho_{Y_1, X_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-0.923\sqrt{5.83}}{\sqrt{5}} = -0.998$$

Notice, here that variable X_2, with coefficient -0.923, receives the greatest weight in component Y_1,

It also has the largest correlation (in absolute value) with Y_1. Then correlation of X_1 with Y_1 0.925 is almost as large as that for X_2, indicating that variables are about equally important to first principal component. Both the coefficients are reasonably large and they have opposite signs, we would argue that both variables aid in interpretation of Y_1.

And there,

$$\rho_{Y_2,X_1} = \rho_{Y_2,X_2} = 0 \quad \text{, and}$$

$$\rho_{Y_2,X_3} = \frac{e_3^1 \sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} \Rightarrow 1$$

Here remaining correlation are neglected as third component is unimportant.

③ EM application :-

Given the model setup:

A paper's true value:

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

A reviewer's bias:

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

A reviewer's score for a given paper:

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

Independency: The variables y^{pr} & z^{pr} are independent
the variable x, y, z for different paper
reviewer pairs are also jointly independent.

a) E-step:-

Hint given from definition of problem:

$$x^{(pr)} = y^{(pr)} + z^{(pr)} + \varepsilon^{(pr)}$$

where $\varepsilon \sim N(0, \sigma^2)$

So x^{pr} follows a normal distributions
that is the sum of multiple independent
normal distributions:

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

For Joint distribution, $P(y^{pr}, z^{pr}, x^{pr})$,

its mean vector.

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

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

$\mathrm{Cov}(A, A+B) = \sigma_A^2$ if A, B are normally distributed random variables and independent.

Therefore, following a trivariate normal distribution.

$$P(x^{pr}, y^{pr}, z^{pr}, \mu_p, \nu_r, \sigma_p^2, \tau_r^2)$$

$$= \frac{1}{(2\pi)^{2/3} |\Sigma_{pr}|^{1/2}} \exp\left[\frac{1}{2} (a^{(pr)} - m_{pr})^T \Sigma_{pr}^{-1} (a^{pr} - m_{pr})\right]$$

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

(ii) let x_1 and x_2 be two multivariate normal random variables,

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

then let x be a new multivariate random variable after stacking x_1 & x_2

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

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

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

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

$$\mu_1 = [\mu_p, \nu_\gamma]^T$$

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

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

$$x_2 = x(p\gamma)$$

$$\mu_2 = \mu_p + \nu_\gamma$$

$$\Sigma_{11} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \tau_\gamma^2 \end{bmatrix}$$

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

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

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

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

$$Q_{pr}(y^{pr}, z^{pr}) = P\left((y^{(pr)}, z^{(pr)}) / x^{pr} \right]$$

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

b) At the E-step, we calculated,

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

Then the lower bound for log likelihood

$$l(\mu_p, \nu_r, \sigma_p^2, \tau_r^2)$$

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

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

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{pr}, z^{pr})} \log \left\{ \frac{\frac{1}{(2\pi)^{4/3} (\Sigma_{pr})^{1/2}} \exp\left[\frac{-1}{2} (a^{pr} - m_{pr})^T \Sigma_{pr}^{-1} (a^{pr} - m_{pr}) \right]}{w(y^{pr}, z^{pr})} \right.$$

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

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

$$m^{pr} = (\mu_{pr}, \nu_r, \mu_p + \nu_r)^T$$

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

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

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

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

Now.

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

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

Setting $\frac{\partial l}{\partial \mu_i}$ to 0 and simplifying we get,

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

Similarly,

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

Setting $\frac{\partial l}{\partial \theta_i^2}$ to 0 and simplifying

we get

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

To update the bias of j^{th} reviewer

ν_j

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

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

Setting $\frac{\partial l}{\partial \nu_j} = 0$ we get

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

To update variance of γ_j i.e

$$\frac{\partial l}{\partial \tau_r^2} = \frac{-1}{2} \sum_{p=1}^{P} \sum_{y,z} w_{(y^{pj}, z^{pj})} \left(\frac{1}{\tau_j^2} - \frac{(z^{pj} - \gamma_j)^2}{\tau_j^4} \right)$$

Now setting $\frac{\partial l}{\partial \tau_r^2} = 0$ and the simplyfying we get

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

Conclusion:-

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

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

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

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

Interpretation of update:-

$$l(\mu_p, \nu_r, \sigma_p^2, \tau_r^2)$$

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

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{pr}, z^{pr})} \Big(\log P(x^{pr} | y^{pr}, z^{pr}) + \log P(y^{pr}, z^{pr}) - \log W(y^{pr}, z^{pr}) \Big)$$