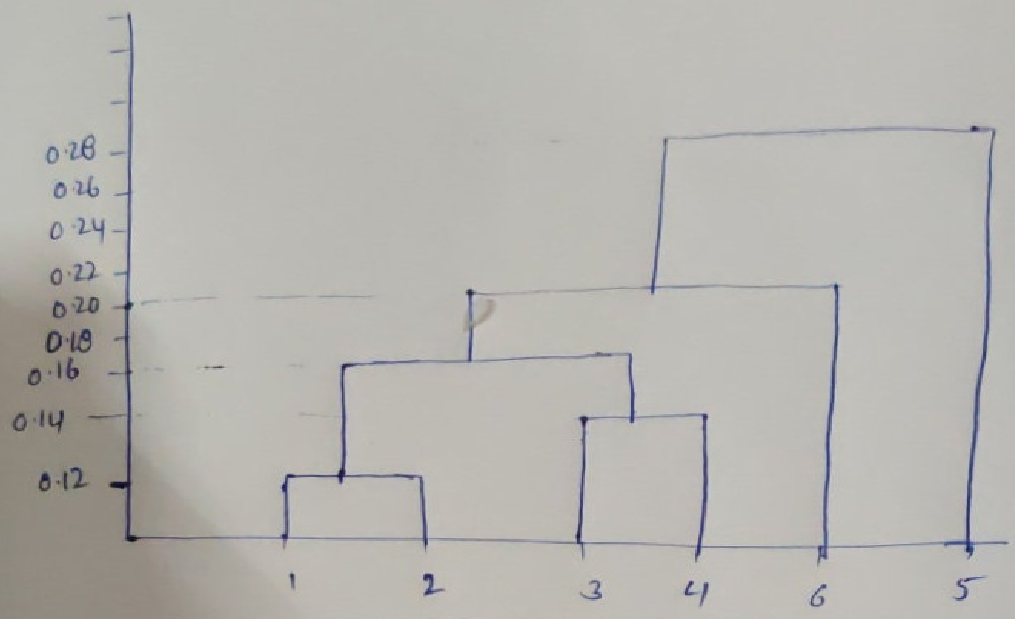
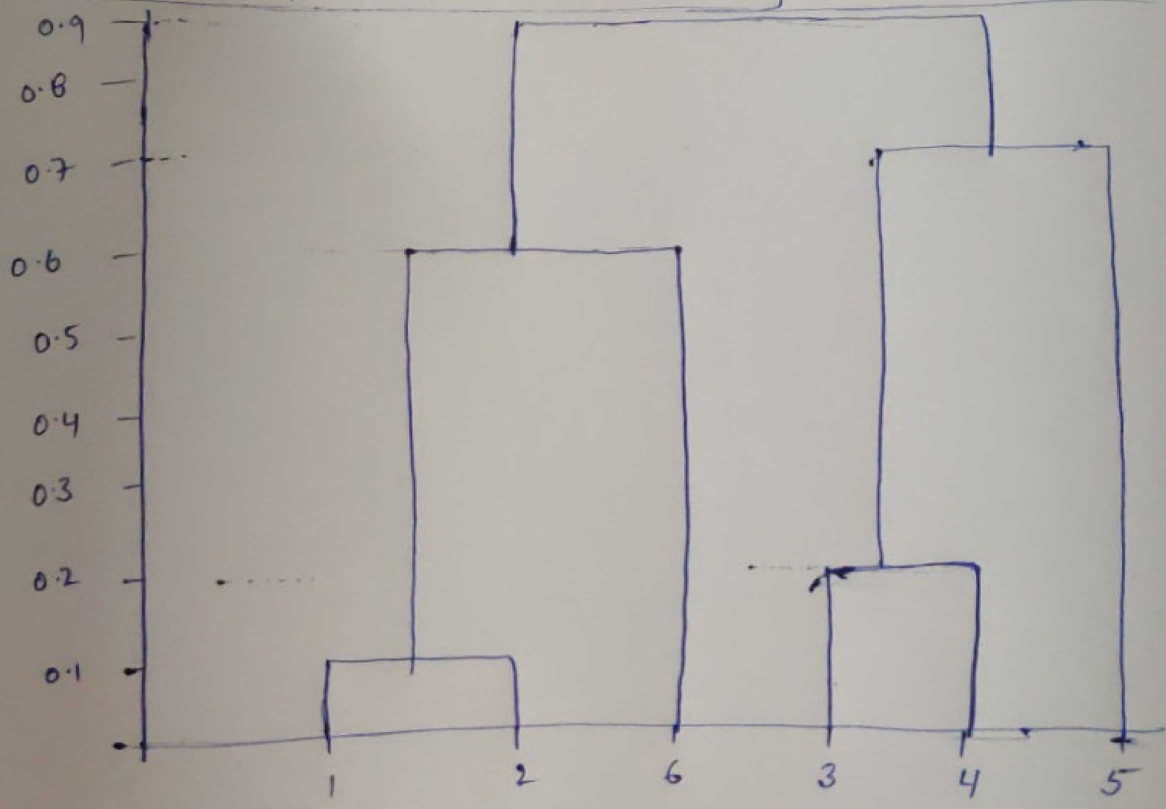
Assigment 5:-

1a Single link: Distance b/w 2 clusters is minimum distance between members of 2 clusters.



Dendogram - Single Link

(1b)



Dendogram with complete link.

15. The first step that the complete link clustering differs from single link is where x_1 x_2 & x_6 are grouped together by dist (x_1, x_2, x_6) = dist (x_2, x_6) = 0.61 we would want dist $(x_1 x_2, x_3 x_4)$ = dist (x_1, x_4) to be smaller than this value such as 0.53.

Then we want dist $(\overset{x_1}{A}\ \overset{x_2}{B}\ \overset{x_3}{C}\overset{x_4}{D}\ \overset{x_6}{F})$ = dist $(\overset{x_3}{C}, \overset{x_6}{F})$ = 0.93. to be smallest so that $\overset{x_1}{A}$ $\overset{x_2}{B}$ $\overset{x_3}{C}$ $\overset{x_4}{D}$ & $\underset{F}{x_6}$ are grouped together. We set this value to 0.63. After these chages. bothe become identical.

2(a) $x' = [x_1, x_2, \dots x_p]$

Covariance matrix Σ

Eigen value, vector pairs $(\lambda_1, \varepsilon_1), (\lambda_2, \varepsilon_2) \dots (\lambda_p, \varepsilon_p)$

where $\lambda_1 \geq \lambda_2 > \lambda_3 \dots \lambda_p \geq 0$

$y_1 = e_1' x \quad y_2 = e_2' x \quad \dots \quad y_p = e_p' x$

are principal components.

Sol:- We know

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

$A = \Sigma$

$\Sigma = P \lambda P'$

λ is diagonal matrix of eigen values and

$P = [e_1, e_2 \dots e_p]$

$P'P = PP' = I$

Therefore

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

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

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

Therefore,

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

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

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

2(b) Given random variables x_1, x_2, x_3 have covariance matrix

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

Eigen values are | Eigen vectors are

$\lambda_1 = 5.83$ | $e_1' = [0.383, -0.924, 0)$

$\lambda_2 = 2.00$ | $e_2' = [0,0,1]$

$\lambda_3 = 0.17$ | $e_3' = [0.924, 0.383, 0]$

(i) 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) Do you think x_3 is a principal component? If so? why?

Sol:- The variable x_3 is a principal component because it is uncorrelated with the other two variables.

iii) Demostrate $Var(Y_i) = \lambda_i$, $i = 1,2,3$ $cov(Y_i, Y_k) = 0$, $i \neq k$?

Sol: $Var(Y_1) = Var(0.383 x_1 - 0.924 x_2)$

$= (0.383)^2 Var(x_1) + (0.924)^2 Var(x_2)$

$+ 2(0.383)(-0.924) cov(x_1, x_2)$

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

$= 5.83 = \lambda_1$

$$\boxed{var(Y_1) = 5.83 = \lambda_1}$$

$$\boxed{var(Y_2) = var(x_3) = 2 = \lambda_2}$$

$Var(Y_3) = var(0.924 x_1 + 0.383 x_2)$

$= (0.924)^2 Var(x_1) + 0.383^2 var(x_2)$

$+ 2(0.924)(0.383) cov(x_1, x_2)$

$= 0.853776 + 0.733 + (-1.415568)$

$= 0.17 = \lambda_3$

$$\boxed{\therefore Var(Y_3) = 0.17 = \lambda_3}$$

To prove

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

$cov(Y_1, Y_2) = cov(0.383 x_1 - 0.924 x_2, x_3)$

$= 0.383\ cov(x_1, x_3)$

$- 0.924\ cov(x_2, x_3)$

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

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

Similarly,

$cov(Y_2, Y_3) = cov(x_3, 0.924 x_1 + 0.383 x_2)$

$= 0.924\ cov(\overset{0}{x_3}, x_1) + 0.383\ cov(\overset{0}{x_3, x_2})$

~~cov(Y_2, Y_3~~

$$\boxed{cov(Y_2, Y_3) = 0}$$

$cov(Y, Y_3) = cov[0.383 x_1 - 0.924 x_2),$
$(0.924 x_1 + 0.383 x_2)]$

$= (0.383)(0.924)\,cov(x_3, x_1) + 0.383^2\, cov(x_1, x_2)$
$- 0.924^2\, cov(x_2, x_1) - 0.924 \times 0.383\, cov(x_2, x_2)$

$= 0.353892(1) + 0.146689(-2)$
$- 0.853776(-2) - 0.353892(5)$

$$\boxed{\therefore cov(Y_1, Y_3) = 0}$$

(IV) Do you think any of principal components could be ignored / eliminated?

Sol: 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 component 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 ~~proportion~~ ~~(5.83+2~~ population variance.

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

$\rho_{Y_1, X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sigma_{11}} = \frac{0.383\sqrt{5.83}}{\sqrt{1}}$
$= 0.925$

$$\rho_{y,x_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-0.924\sqrt{5.83}}{\sqrt{5}}$$

$$= -0.998$$

Notice here that Variable x_2, with Coefficient -0.924, receives the greatest weight in component y_1,

It also has the largest correlation (in absolute Value) with y_1, The 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 here,

$$\rho_{y_2,x_1} = \rho_{y_2,x_2} = 0 \text{ and}$$

$$\rho_{y_2,x_3} = \frac{e_3'\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} \Rightarrow 1$$

Here remaining correlation are neglected as third component is unimportant.

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:

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

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 jointy independent.

(a) E-step:-

Hint given from definition of problem:

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

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

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.

$$M_{pr} = [u_p, v_r, u_p + v_r]^T$$

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

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

Therefore, following a trivariate normal distributions

$$P(x^{pr}, y^{pr}, z^{pr}; u_p, v_r, \sigma_p^2 \tau_r^2]$$

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

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

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

$$x_1 \sim N(u_1, \Sigma_{11})$$

$$x_2 \sim N(u_2, \Sigma_{22})$$

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

$$x = [x_1, x_2]^T$$

$$\sim N[(\mu_1, \mu_2)^T, \Sigma]$$

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

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

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

$$\mu_1 = [\mu_p, \nu_r]^T.$$

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

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

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

$$\mu_2 = \mu_p + \nu_r.$$

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

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

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

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

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

Scanned by Scanner Go

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

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

(b) At the $\mathcal{E}$-step, we calculated,

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

Then the lower bound for log likelihood $l(\mu_p, v_r, \sigma_p^2, \tau_r^2)$

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

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

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

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

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

$$m_{pr} = (\mu_p, v_r, \mu_p + v_r)^T$$

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

$$|\Sigma pr| = \sigma_p^2 \, \tau_r^2 \, \sigma^2$$

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

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

Now

$$\frac{\partial l}{\partial u_1} = \sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{ir}, z^{ir})} \left[\frac{1}{\sigma_1^2}, 0, \frac{-2}{\sigma^2}\right] \begin{bmatrix} y^{ir} \\ z^{ir} \\ z^{ir} \end{bmatrix}$$

$$\sum_{r=1}^{R} \sum_{(y,z)} w_{y^{ir}, z^{ir}} \left[\frac{1}{\sigma_1^2}, 0, \frac{-2}{\sigma_i^2}\right] \begin{bmatrix} \mu_1 \\ \gamma_r \\ \mu_i + \nu_r \end{bmatrix}$$

Setting $\frac{\partial l}{\partial u_1}$ to 0 and simplifying we get,

$$\boxed{u_1 = \frac{\sum_{r=1}^{R} \sum_{y,z} w_{(y^{ir}, z^{ir})} \left(\frac{y^{ir}}{\sigma_i^2} - \frac{2(z^{ir} - \nu_r)}{\sigma^2}\right)}{\sum_{r=1}^{R} \sum_{y,z} w_{(y^{ir}, z^{ir})} \left(\frac{1}{\sigma_1^2} - \frac{2}{\sigma^2}\right)}}$$

Similarly

$$\frac{\partial l}{\partial \varepsilon_1^2} = \frac{-1}{2} \sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{ir}, z^{ir})} \left(\frac{1}{\sigma_1^2} - \frac{(y^{ir} - u_1)^2}{\sigma_1^4}\right)$$

Setting $\frac{\partial l}{\partial \sigma_1^2}$ to 0 and simplyfying.

we get,

$$\sigma_1^2 = \frac{\sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{ir}, z^{ir})} (y^{ir} - \mu_i)^2}{\sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{ir}, z^{(ir)})}}$$

- - - - - - - - - - - - - - - - - - -

To update the bias of j^{th} reviewer. ν_i r_i

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

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

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

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

To update variance

$$\frac{\partial l}{\partial \gamma_r^2} = -\frac{1}{2} \sum_{p=1}^{P} \sum_{(y,z)} w_{(y^{pj}, z^{pj})} \left(\frac{1}{\gamma_j^2} - \frac{(z^{pj} - r_1)^2}{\gamma_1^4}\right)$$

Now setting $\frac{\partial l}{\partial \gamma_r^2} = 0$ and the

Simplyfying we get

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

Scanned by Scanner Go

Conclusion:

$$\mu_i = \frac{\sum_{r=1}^{R} \sum_{y,z} w_{(y^{ir}, z^{ir})} \left(\frac{y^{i2}}{\sigma_i^2} - 2\frac{(x^{ir} - v_r)}{\sigma^2} \right)}{\sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{ir}, z^{ir})} \left(\frac{1}{\sigma_i^2} - \frac{2}{\sigma^2} \right)}$$

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

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

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

Interpretation of update:

$$l(\mu_p, v_r, \sigma_p^r, v_r^2)$$

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

$$= \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{(y,z)} w_{(y^{pr}, z^{pr})} \left(\log P(x^{pr} / y^{pr}, z^{pr})_{+} \right.$$

$$\left. \log P(y^{pr}, z^{pr}) - \log w_{(y^{pr}, z^{pr})} \right)$$