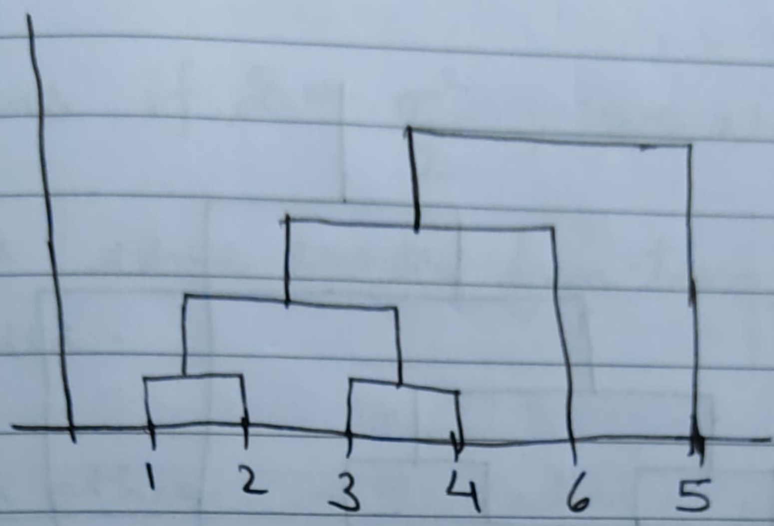
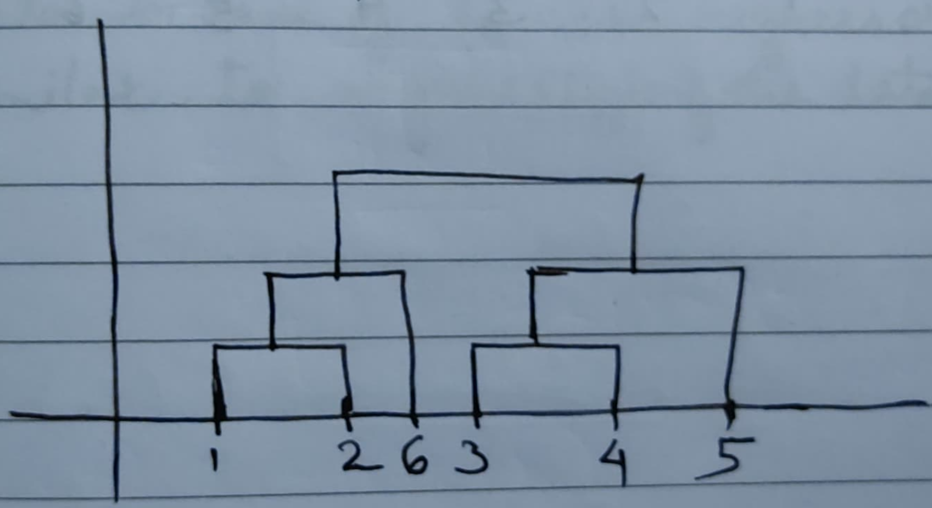
AI20BTECH11011

Assignment 5.

Q.1.) a.)



b.)



c.) If we change $x_1 - x_4$ distance to 0.60 and $x_3 - x_6$ distance to 0.6[illegible]6, the dendrograms to the above two questions will end up being the same.

Q.2.) a) Assumptions: $X_1, X_2, \dots X_p$ are zero mean.

$$\Rightarrow \Sigma = E\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}^T [X_1\ X_2 \dots X_p]\right] = E[X X^T]$$

$$= \begin{bmatrix} E[X_1^T X_1] & E[X_1^T X_2] & \cdots & E[X_1^T X_p] \\ E[X_2^T X_1] & E[X_2^T X_2] & \cdots & E[X_2^T X_p] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_p^T X_1] & E[X_p^T X_2] & \cdots & E[X_p^T X_p] \end{bmatrix}$$

$$\Rightarrow \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = E[X_1^T X_1] + E[X_2^T X_2] + \dots + E[X_p^T X_p].$$

$$Var(X_i) = E[(X_i - \mu_i)^T (X_i - \mu_i)]$$

$$= E[X_i^T X_i].$$

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

Since Σ is square symmetric, we can use spectral decomposition theorem to say

$$\Sigma = V D V^T, \quad V = [e_1\ e_2 \dots e_p]$$

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

where V is orthogonal.

$\Rightarrow Tr(\Sigma) = Tr(VDV^T)$.

Using $Tr(ABC) = Tr(CAB)$,

$Tr(\Sigma) = Tr(V^TVD) = Tr(D)$

$$\boxed{\Rightarrow \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p}$$

$Var(Y_i) = E[(Y_i - \eta_i)(Y_i - \eta_i)^T]$

Since $X_1, X_2, \dots, X_p$ are zero mean & Y_i is linear combination of $X_1^T, X_2^T, \dots, X_p^T$, Y_i is also zero mean.

$\Rightarrow Var(Y_i) = E[Y_i Y_i^T]$

$= E[e_i^T X X^T e_i]$

$= \cancel{E[e_i^T} \; e_i^T E[XX^T] e_i$

$= e_i^T \Sigma e_i$

$= e_i^T \lambda_i e_i$

$= \lambda_i$ ($\because e_i$s are unit norm)

$$\boxed{\Rightarrow \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i)}.$$

b.) i.) $Y_1 = e_1^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \cancel{X_1 - 2X_2}$

$= 0.383 X_1 - 0.924 X_2$

$Y_2 = e_2^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = X_3$

$Y_3 = e_3^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = 0.924 X_1 + 0.383 X_2$

ii.) X_3 is a principal component because from Σ, we can infer $X_1^T X_3 = X_2^T X_3 = 0$.

$\Rightarrow$ X_3 is linearly independent of X_1 & X_2.

Thus X_3 must be a principal component.

iii.) $Var(Y_1) = Y_1 Y_1^T$

$= e_1^T \Sigma e_1$

$= \begin{bmatrix} 0.383 & -0.924 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0.383 \\ -0.924 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} 0.383 & -0.924 & 0 \end{bmatrix} \begin{bmatrix} 2.231 \\ -5.386 \\ 0 \end{bmatrix}$

$= 5.83 = \lambda_1$

$$\mathrm{Var}(Y_2) = e_2^T \Sigma e_2$$

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

$$= 2 = \lambda_2$$

$$\mathrm{Var}(Y_3) = e_3^T \Sigma e_3$$

$$= \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0.924 \\ 0.383 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 0.158 \\ 0.067 \\ 0 \end{bmatrix}$$

$$= 0.17 = \lambda_3$$

$$\mathrm{Cov}(Y_1, Y_2) = e_1^T \Sigma e_2$$

$$= \begin{bmatrix} 0.383 & -0.924 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.383 & -0.924 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

$$= 0$$

$$\mathrm{Cov}(Y_3, Y_2) = e_3^T \Sigma e_2$$

$$= \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

$$= 0$$

$$\mathrm{Cov}(Y_3, Y_1) = \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0.383 \\ -0.924 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix} \begin{bmatrix} 2.231 \\ -5.386 \\ 0 \end{bmatrix}$$

$$= 0.$$

Hence $\mathrm{Var}(Y_i) = \lambda_i$

$\mathrm{Cov}(Y_i, Y_k) = 0$ if $i \neq k$.

iv) The third principal component Y_3 can be eliminated because the variance along Y_3 is 0.17 which is much less than the variances along Y_1 (5.83) & Y_2 (2.00).

Q.3) a) i) Since $x^{pr} \sim N(y^{pr} + z^{pr}, \sigma^2)$,

we can write $x^{pr} = y^{pr} + z^{pr} + \epsilon$,
where $\epsilon \sim N(0, \sigma^2)$.

$$E\left[\begin{pmatrix} y^{pr} \\ z^{pr} \\ x^{pr} \end{pmatrix}\right] = \begin{pmatrix} \mu_p \\ \nu_r \\ \mu_p + \nu_r \end{pmatrix}.$$

$$Var(y^{pr}) = \sigma_p^2$$
$$Var(z^{pr}) = \tau_r^2$$

$$Var(x^{pr}) = Var(y^{pr}) + Var(z^{pr}) + Var(\epsilon).$$
$$= \sigma_p^2 + \tau_r^2 + \sigma^2$$

$Cov(y^{pr}, z^{pr}) = 0$ $\quad$ ($\because$ y^{pr} & z^{pr} are independant).

$$Cov(y^{pr}, x^{pr}) = Cov(y^{pr}, y^{pr}) + Cov(z^{pr}, y^{pr}) + Cov(y^{pr}, \epsilon)$$

$$= \sigma_p^2$$

$$Cov(z^{pr}, x^{pr}) = Cov(y^{pr}, z^{pr}) + Cov(z^{pr}, z^{pr}) + Cov(z^{pr}, \epsilon)$$

$$= \tau_r^2$$

$$\Rightarrow \Sigma = \begin{bmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_\gamma^2 & \tau_\gamma^2 \\ \sigma_p^2 & \tau_\gamma^2 & \sigma^2+\sigma_p^2+\tau_\gamma^2 \end{bmatrix}$$

ii.) Rules for conditioning on subsets of jointly gaussian random variables:

If $Y \sim N(\mu, \Sigma)$, consider partitioning μ & Y into

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

with a similar partition of Σ into

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

Then $(y_1 \mid y_2) \sim N(\bar{\mu}, \bar{\Sigma})$

$$\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)$$

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Hence, $y_1 \equiv (y^{p\gamma}, z^{p\gamma})$.

$y_2 \equiv (x^{p\gamma})$

$$\Rightarrow \bar{\mu} = \begin{bmatrix} \mu_p \\ \nu_\gamma \end{bmatrix} + \begin{bmatrix} \sigma_p^2 \\ \tau_\gamma^2 \end{bmatrix} \cdot \frac{x^{p\gamma} - \mu_p - \nu_\gamma}{\sigma^2 + \sigma_p^2 + \tau_\gamma^2}$$

$$\bar{\Sigma} = \begin{bmatrix} \sigma_\rho^2 & 0 \\ 0 & \tau_\gamma^2 \end{bmatrix} - \begin{bmatrix} \sigma_\rho^2 \\ \tau_\gamma^2 \end{bmatrix} \times \frac{1}{\sigma_\rho^2 + \tau_\gamma^2 + \sigma^2} \times \begin{bmatrix} \sigma_\rho^2 & \tau_\gamma^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_\rho^2 & 0 \\ 0 & \tau_\gamma^2 \end{bmatrix} - \frac{1}{\sigma_\rho^2 + \tau_\gamma^2 + \sigma^2} \begin{bmatrix} \sigma_\rho^4 & \sigma_\rho^2 \tau_\gamma^2 \\ \sigma_\rho^2 \tau_\gamma^2 & \tau_\gamma^4 \end{bmatrix}.$$

Thus $p(y^{\rho\gamma}, z^{\rho\gamma} \mid x^{\rho\gamma}) = \mathcal{N}(\bar{\mu}, \bar{\Sigma})$

$$= \frac{\exp\left(\frac{-1}{2} (X - \bar{\mu})^T \bar{\Sigma}^{-1} (X - \bar{\mu}) \right)}{\sqrt{(2\pi)^2 |\bar{\Sigma}|}}$$

where $X = \begin{bmatrix} y^{\rho\gamma} \\ z^{\rho\gamma} \end{bmatrix}$, $|\bar{\Sigma}| = \det(\bar{\Sigma})$.

Q.3b) Let $E_q[\cdot]$ denote expectation w.r.t. $q^{pr} \equiv (y^{pr}, z^{pr})$ drawn from a distribution with density $p(y^{pr}, z^{pr} \mid x^{pr})$ as derived in part (a) of the question.

Thus the likelihood function we want to maximize is

$$l = \sum_{p,r} E_q[\ln p(\epsilon, y^{pr}, z^{pr})]$$

$$= \sum_{p,r} E_q[\ln(p(y^{pr}, z^{pr}))]$$

(Since ϵ, y^{pr} & z^{pr} are independent)

$$= \sum_{p,r} E_q\left[\ln\frac{1}{2\pi\sigma_p\tau_r} - \frac{(y^{pr}-\mu_p)^2}{2\sigma_p^2} - \frac{(z^{pr}-\nu_r)^2}{2\tau_r^2}\right]$$

$$= \sum_{p,r} E_q\left[\ln\frac{1}{2\pi\sigma_p\tau_r} - \frac{((y^{pr})^2+\mu_p^2-2y^{pr}\mu_p)}{2\sigma_p^2} - \frac{((z^{pr})^2+\nu_r^2-2z^{pr}\nu_r)}{2\tau_r^2}\right]$$

$$= \sum_{p,r} \ln\frac{1}{2\pi\sigma_p\tau_r} - \frac{(E_q[(y^{pr})^2]+\mu_p^2-2E_q[y^{pr}]\mu_p)}{2\sigma_p^2} - \frac{(E_q[(z^{pr})^2]+\nu_r^2-2E_q[z^{pr}]\nu_r)}{2\tau_r^2}$$

We already ~~the~~ know the ~~conditi~~ distribution $P(y^{pr}, z^{pr} | x^{pr}) \equiv Q(y^{pr}, z^{pr})$

Thus we can find $E_Q[y^{pr}]$

$$= E\left[\int_{-\infty}^{\infty} Q(y^{pr}, z^{pr})\, dz^{pr}\right]$$

Let $\mu_{pr} := E_Q[y^{pr}]$.

Likewise, we find $\sigma_{pr}^2 := E_Q[(y^{pr})^2] - \mu_{pr}^2$

$$\nu_{pr} := E_Q[z^{pr}]$$

$$\tau_{pr}^2 := E_Q[(z^{pr})^2] - \nu_{pr}^2$$

$$\Rightarrow l = \sum_{p,r} \ln \frac{1}{2\pi\sigma_p\tau_r} - \frac{(\sigma_{pr}^2 + \mu_{pr}^2 + \mu_p^2 - 2\mu_{pr}\cdot\mu_p)}{2\sigma_p^2}$$

$$- \frac{(\tau_{pr}^2 + \nu_{pr}^2 + \nu_r^2 - 2\nu_{pr}\cdot\nu_r)}{2\tau_r^2}$$

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

$$\sum_{r=1}^{R} \frac{(\mu_p - \mu_{pr})}{\sigma_p^2} = 0 \Rightarrow \mu_p = \frac{1}{R}\sum_{r=1}^{R} \mu_{pr}$$

Likewise, setting $\frac{\partial l}{\partial \nu_r} = 0$, we get

$$\nu_p = \frac{1}{p} \sum_{p=1}^{p} \nu_{pr}$$

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

$$\sum_{r=1}^{R} \left[\frac{-1}{\sigma_p} + \frac{1}{\sigma_p^3} \left(\sigma_{pr}^2 + \mu_{pr}^2 - 2\mu_{pr} \cdot \mu_p + \mu_p^2 \right) \right] = 0$$

$$\Rightarrow \sigma_p^2 = \frac{1}{R} \sum_{r=1}^{R} \left(\sigma_{pr}^2 + \mu_{pr}^2 + \mu_p^2 - 2\mu_{pr} \cdot \mu_p \right)$$

Likewise,

$$\cdot \tau_r^2 = \frac{1}{P} \sum_{p=1}^{P} \left(\tau_{pr}^2 + \nu_{pr}^2 + \nu_r^2 - 2\nu_{pr} \cdot \nu_r \right)$$