Q1.

(a) True

(b) Suppose these three groups x_a, x_b, x_c are jointly Gaussian, then we can define:

$$p(x_a, x_b, x_c) = \mathcal{N}(x \mid \mu, \Sigma)$$

where $\mu = \begin{bmatrix} \mu_a \\ \mu_b \\ \mu_c \end{bmatrix}$ $\quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{bmatrix}$

According to the marginalization of Gaussian distribution, if x_c have been marginalized out, we have

$$p(x_a, x_b) = \mathcal{N}(x \mid \mu, \Sigma)$$

where $\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}$ $\quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$

Let's define $\Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$

Since $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix}$ $\quad M = (A - BD^{-1}C)^{-1}$

then we have $\begin{cases} \Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab} \cdot \Sigma_{bb}^{-1} \Sigma_{ba})^{-1} \\ \Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1} \Sigma_{ab} \Sigma_{bb}^{-1} \end{cases}$

Thus, we have $p(x_a \mid x_b) = \mathcal{N}(x \mid \mu_{a|b}, \Lambda_{aa}^{-1})$

where $\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (x_b - \mu_b)$

Q2

(a) Given the properties of Gaussian distribution.

$$p(z) = \mathcal{N}(z \mid \mu, \Sigma)$$

where $\mu = \begin{bmatrix} \mu_x \\ \end{bmatrix} = \begin{bmatrix} \mu \\ \end{bmatrix}$ $\quad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \end{bmatrix} = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1} A^T \\ \end{bmatrix}$

$$\left[\mu_y\right] \; \left[A\mu+b\right] \qquad \left[\Sigma_{yx} \quad \Sigma_{yy}\right] \; \left[A\Lambda^{-1} \quad L^{-1}+A\Lambda^{-1}A^T\right]$$

and $p(x) = N(x \mid \mu_x, \Sigma_{xx})$

Thus, we have $p(x) = N(x \mid \mu, \Lambda^{-1})$

(b) Using the properties of Gaussian distribution,

we have $p(y \mid x) = N(y \mid \mu_{y|x}, \Lambda_{yy}^{-1})$

where $\mu_{y|x} = \mu_y + \Sigma_{yx} \cdot \Sigma_{xx}^{-1} (x - \mu_x)$

$\Lambda_{yy} = (\Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy})^{-1}$

Thus, $\mu_{y|x} = A\mu + b - A\Lambda^{-1} \cdot \Lambda \cdot (x - \mu) = Ax + b$

$\Lambda_{yy}^{-1} = L^{-1} + A\Lambda^{-1}A^T - A\Lambda^{-1}\Lambda\Lambda^{-1}A^T = L^{-1}$

$\Rightarrow p(y \mid x) = N(y \mid Ax + b, L^{-1})$

Q3 Let's denote the log likelihood function as

$$L(\Sigma^{-1}) = \ln p(X \mid \mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) + \frac{N}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)^T \cdot \Sigma^{-1} \cdot (x_n - \mu)$$

Taking the derivative of $L(\Sigma)$ with respect to Σ^{-1}, we have

$$\frac{\partial}{\partial \Sigma^{-1}} L(\Sigma^{-1}) = \frac{N}{2} \cdot \frac{\partial}{\partial \Sigma^{-1}} \ln|\Sigma^{-1}| - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}} \sum_{i=1}^{N}(x_n - \mu)^T \cdot \Sigma^{-1}(x_n - \mu)$$

Since
① $(x_n - \mu)^T \cdot \Sigma^{-1}(x_n - \mu)$ is a scalar
② $tr(c) = c$ if c is a scalar
③ $tr(AB) = tr(BA)$
④ $\frac{\partial}{\partial A} tr(A) = I$, $\frac{\partial}{\partial A} \ln|A| = (A^{-1})^T$

then we have

$$\frac{\partial}{\partial \Sigma^{-1}} L(\Sigma^{-1}) = \frac{N}{2} \cdot \Sigma^T - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}} \sum_{i=1}^{N} tr((x_n-\mu)^T (x_n-\mu) \cdot \Sigma^{-1})$$

$$= \frac{N}{2}\Sigma^T - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}} tr\left[\sum_{i=1}^{N}(x_n-\mu)(x_n-\mu)^T \cdot \Sigma^{-1}\right]$$

$$= \frac{N}{2}\Sigma^T - \frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)(x_n-\mu)^T$$

$$= \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(x_n-\mu)(x_n-\mu)^T \quad (\text{Gaussian}: \Sigma^T = \Sigma)$$

$$= 0$$

$$\Rightarrow \Sigma = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_n - \mu)(x_n - \mu)^T$$

Since the sample variance is non-singular, then the final result is symmetric and positive definite.

Q5 Using Bayes' theorem, we have

$$p(\mu|X) = \frac{p(X|\mu) \cdot p(\mu)}{p(X)}$$

① Since $X = \{x_1, x_2, \cdots, x_N\}$, and for each x_i, $x_i \sim N(x_i|\mu, \Sigma)$

then we have $p(X|\mu) = \prod_{i=1}^{N} p(x_i|\mu)$

$$= \exp\left(\sum_{i=1}^{n} \ln N(\mu, \Sigma)\right)$$

$$\propto \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \cdot \Sigma^{-1} (x_i - \mu)\right)$$

② Since $p(\mu) = N(\mu | \mu_0, \Sigma_0)$, then

$$p(\mu) = \exp(\ln p(\mu))$$

$$\propto \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\mu - \mu_0)^T \cdot \Sigma_0^{-1} (\mu - \mu_0)\right)$$

③ Since X is given as observation, then $p(X)$ is fixed

Thus, we have:

$$p(\mu|X) \propto p(X|\mu) \cdot p(\mu)$$

$$= \exp\left(-\frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) - \frac{1}{2} \sum_{i=1}^{n} (\mu - \mu_0)^T \Sigma_0^{-1} (\mu - \mu_0)\right)$$

$$\propto \exp\left[-\frac{1}{2}\left[\mu^T (\Sigma_0^{-1} + N\Sigma^{-1}) \cdot \mu - 2\mu^T (\Sigma_0^{-1}\mu_0 + \Sigma^{-1} \sum_{i=1}^{N} x_i)\right]\right]$$

Comparing this expression with the standard form of a Gaussian distribution, we have

$$\mu_{post} = (\Sigma_0^{-1} + N\Sigma^{-1})^{-1} (\Sigma_0^{-1}\mu_0 + \Sigma^{-1} \sum_{i=1}^{N} x_i)$$

$$\Sigma_{post} = (\Sigma_0^{-1} + N\Sigma^{-1})^{-1}$$