1. b) $P_N(n|\vartheta) = \binom{N}{n} \vartheta^n (1-\vartheta)^{N-n}$

$$P_M(m|\vartheta) = \binom{M}{m} \vartheta^m (1-\vartheta)^{M-m}$$

$$\prod_{n=1}^{N_k} \left(\binom{M}{m_{nk}} \vartheta_k^{m_{nk}} (1-\vartheta_k)^{M-m_{nk}} \right) =$$

$$= \left(\prod_{n=1}^{N_k} \binom{M}{m_{nk}} \right) \cdot \vartheta_k^{\sum_{n=1}^{N_k} m_{nk}} \cdot (1-\vartheta_k)^{\sum_{n=1}^{N_k} (M-m_{nk})} \quad \underline{\underline{\log}}$$

$$= \log \left(\prod_{n=1}^{N_k} \binom{M}{m_{nk}} \right) + \sum_{n=1}^{N_k} m_{nk} \log \vartheta_k + \sum_{n=1}^{N_k} (M-m_{nk}) \log(1-\vartheta_k) =$$

$$= \sum_{n=1}^{N_k} \log \binom{M}{m_{nk}} + \sum_{n=1}^{N_k} m_{nk} \log \vartheta_k + \sum_{n=1}^{N_k} (M-m_{nk}) \cdot \log(1-\vartheta_k)$$

$$\hat{\vartheta}_k = \underset{\vartheta_k}{\operatorname{argmax}} \{ L(\vartheta_k) \}$$

$$\frac{dL(\vartheta_k)}{d\vartheta_k} = \frac{\sum_{n=1}^{N_k} m_{nk}}{\vartheta_k} + \frac{\sum_{n=1}^{N_k} (M-m_{nk})}{1-\vartheta_k} (-1) = 0$$

$$\left(\sum_{n=1}^{N_k} m_{nk} \right) (1-\vartheta_k) - \left(\sum_{n=1}^{N_k} (M - m_{nk}) \right) \cdot \vartheta_k = 0$$

$$\sum_{n=1}^{N_k} m_{nk} - \cancel{\vartheta_k \sum_{n=1}^{N_k} m_{nk}} - N_k \cdot M \cdot \vartheta_k + \cancel{\vartheta_k \cdot \sum_{n=1}^{N_k} m_{nk}} = 0$$

$$\vartheta_k = \frac{\sum_{n=1}^{N_k} m_{nk}}{N_k \cdot M} = \frac{\cancel{N_k} \cdot \overline{m}_k}{\cancel{N_k} \cdot M} = \boxed{\frac{\overline{m}_k}{M}}$$

c) $$p(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

$$p(m|\lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$$

$$\prod_{n=1}^{N_k} \frac{\lambda^{m_{nk}}}{m_{nk}!} e^{-\lambda} = \frac{\lambda^{\sum_{n=1}^{N_k} m_{nk}}}{\prod_{n=1}^{N_k} m_{nk}!} e^{-N_k \lambda} \quad \underline{\log}$$

$$\sum_{n=1}^{N_k} m_{nk} \cdot \log \lambda - \log\left(\prod_{n=1}^{N_k} m_{nk}!\right) - N_k \cdot \lambda \cdot \log(e) =$$

$$= \left(\sum_{n=1}^{N_k} m_{nk}\right) \cdot \log \lambda - \sum_{n=1}^{N_k} \cdot \log(m_{nk}!) - N_k \cdot \lambda$$

$$\hat{\lambda} = \arg\max_{\lambda} \{ L(\lambda) \}$$

$$\frac{dL(\lambda)}{d\lambda} = \frac{\sum_{n=1}^{N_k} m_{nk}}{\lambda} - N_k = 0$$

$$\sum_{n=1}^{N_k} m_{nk} - N_k \cdot \lambda = 0$$

$$\hat{\lambda} = \frac{\sum_{n=1}^{N_k} m_{nk}}{N_k} = \overline{m_k}$$

③ a) full class specific covariance matrix

$$p(x|k,\mu_k,\Sigma_k) = \frac{1}{\sqrt{(2\pi)^D \cdot \det(\Sigma_k)}} \cdot \exp\left[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right]$$

$$\prod_{n=1}^{N_k} \frac{1}{\sqrt{(2\pi)^D \cdot \det(\Sigma_k)}} \cdot \exp\left[-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right] \xrightarrow{\log}$$

$$\Rightarrow \sum_{n=1}^{N_k}\left(-\log\sqrt{(2\pi)^D} - \log\left(\sqrt{\det(\Sigma_k)}\right) + \log\left(\exp\left[-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\right]\right)\right) =$$

$$= \sum_{n=1}^{N_k}\left(-\log\sqrt{(2\pi)^D} - \underbrace{\log\left(\frac{1}{\sqrt{\det(\Sigma_k^{-1})}}\right)}_{1^\circ} \underbrace{- \frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}_{2^\circ}\right) =$$

* FOR THE PART 1° I will USE THE PROVIDED FORMULA

$$\frac{d\det(A)}{dA_{ij}} = (A^{-1})_{ij} \cdot \det A \Rightarrow \frac{d\det(A)}{dA} = A^{(-1)} \cdot \det(A)$$

* FOR PART 2° I will USE THE CHARACTERISTICS OF THE TRACE OF A SQUERE MATRIX

$$\boxed{tr[A] = \sum_i A_{ii}}$$

AN IMPORTANT PROPERTY OF TRACE IS ITS INVARIANCE UNDER CYCLYCAL PERMUTATIONS

$$\boxed{tr[ABC] = tr[CAB] = tr[BCA]}$$

ALSO ONE PROPERTY WHICH WILL HELP US IS for vector x AND MATRIX A

$$\underbrace{x^T A x}_{\text{SCALAR}} = tr[x^T A x] = tr(x x^T A)$$

SCALAR ⇒ AND $tr(a) = a$

$$\frac{d}{dA_{ij}} tr[AB] = \frac{d}{dA_{ij}} \sum_k \sum_e A_{ke}B_{ek} = B_{ji}$$

$$\Rightarrow \frac{d}{dA} tr[BA] = \frac{d}{dA} tr[AB] = B^T$$

$$\frac{d}{dA} x^T A x = \frac{d}{dA} tr[x^T A x] = \frac{d}{dA} tr[x x^T A] =$$

$$= [x x^T]^T = x x^T \quad (i)$$

$\hookrightarrow$ SINCE X IS A VECTOR

$$\hat{\Sigma}_k = \underset{\Sigma_k}{\operatorname{argmax}} \{L(\Sigma_k)\}$$

$$\frac{dL(\Sigma_k)}{d\Sigma_k^{-1}} = \sum_{n=1}^{N_k} \left(- \frac{\log\left(\frac{1}{\sqrt{\det(\Sigma^{-1})}}\right)}{d\Sigma_k^{-1}} - \frac{\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{(-1)} (x_n - \mu_k)}{d\Sigma_k^{-1}} \right) =$$

$$= \sum_{n=1}^{N_k} \left(\frac{1}{\cancel{\sqrt{\det(\Sigma_k^{-1})}}} \cdot \frac{1}{2\cancel{\sqrt{\det(\Sigma_k^{-1})}}} \cdot \Sigma_k \cdot \cancel{\det(\Sigma_k^{-1})} - \frac{1}{2} \frac{tr[(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)]}{d\Sigma_k^{-1}} \right) =$$

$$= \frac{N_k}{2} \cdot \Sigma_k - \frac{1}{2} \sum_{n=1}^{N_k} \frac{tr[(x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1}]}{d\Sigma_k^{-1}} = \text{from } (i)$$

$$= \frac{N_k}{2} \cdot \Sigma_k - \frac{1}{2} \sum_{n=1}^{N_k} (x_n - \mu_k)(x_n - \mu_k)^T = 0$$

$$\boxed{\hat{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} (x_n - \mu_k)(x_n - \mu_k)^T}$$

b) diagonal class specific covariance MATRIX

$$p(x \mid k, \mu_k, \sigma_k^2) = \frac{1}{\prod_{d=1}^{D} \sqrt{2\pi \sigma_{kd}^2}} \exp\left[-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_d - \mu_{kd}}{\sigma_{kd}}\right)^2\right]$$

$$\prod_{n=1}^{N_k} p(x_n \mid k, \mu_k, \sigma_k^2) = \prod_{n=1}^{N_k} \frac{1}{\prod_{d=1}^{D} \sqrt{2\pi \sigma_{kd}^2}} \exp\left[-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_{nd} - \mu_{kd}}{\sigma_{kd}}\right)^2\right]$$

$$\xRightarrow{\log} \sum_{n=1}^{N_k}\left(\log \frac{1}{\prod_{d=1}^{D}\sqrt{2\pi \sigma_{kd}^2}} + \log\left|\exp\left[-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_{nd}-\mu_{kd}}{\sigma_{kd}}\right)^2\right]\right|\right) =$$

$$= \sum_{n=1}^{N_k}\left(-D \cdot \log\sqrt{2\pi} - \sum_{d=1}^{D}\log \sigma_{kd} - \frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_{nd}-\mu_{kd}}{\sigma_{kd}}\right)^2\right)$$

$$\hat{\sigma}_{kd} = \underset{\sigma_{kd}}{\arg\max}\left(L(\sigma_{kd})\right)$$ for specific class k and dimension d

$$\frac{dL(\sigma_{kd})}{d\sigma_{kd}} = \sum_{n=1}^{N_k}\left(-\frac{1}{\sigma_{kd}} - \frac{1}{2}\cdot(x_{nd}-\mu_{kd})^2 \cdot \left(-\frac{2}{\sigma_{kd}^3}\right)\right) = 0 \Big/ \cdot \sigma_{kd}^3$$

$$\sum_{n=1}^{N_k}\left(-\sigma_{kd}^2 + (x_{nd}-\mu_{kd})^2\right) = 0$$

$$N_k \cdot \sigma_{kd}^2 = \sum_{n=1}^{N_k}(x_{nd}-\mu_{kd})^2$$

$$\boxed{\hat{\sigma}_{kd}^2 = \frac{1}{N_k}\cdot\sum_{n=1}^{N_k}(x_{nd}-\mu_{kd})^2}$$

c) pooled covariance matrix

$$p(x|\mu_k, \Sigma) = \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left[-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)\right]$$

$$\prod_{n=1}^{N} p(x|\mu_k, \Sigma) = \prod_{n=1}^{N} \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left[-\frac{1}{2}(x_n-\mu_{kn})^T \Sigma^{-1} (x_n-\mu_{kn})\right] =$$

$$\xrightarrow{\log} \sum_{n=1}^{N} \left(-\log\sqrt{(2\pi)^D} - \log(\sqrt{\det(\Sigma)}) - \frac{1}{2}(x_n-\mu_{kn})^T \Sigma^{-1} (x_n-\mu_{kn})\right) =$$

$$= \sum_{n=1}^{N} \left(-\log\sqrt{(2\pi)^D} - \log\left(\frac{1}{\sqrt{\det(\Sigma^{-1})}}\right) - \frac{1}{2}(x_n-\mu_{kn})^T \Sigma^{-1} (x_n-\mu_{kn})\right)$$

* WITH THE SAME ARGUMENTS AS PART A $\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmax}}(L(\Sigma))$

$$\Rightarrow \frac{dL(\Sigma)}{d\Sigma^{-1}} = \sum_{n=1}^{N}\left(-\frac{\log\frac{1}{\sqrt{\det(\Sigma^{-1})}}}{d\Sigma^{-1}} - \frac{1}{2}\frac{(x_n-\mu_{kn})^T \Sigma^{-1}(x_n-\mu_{kn})}{d\Sigma^{-1}}\right) =$$

$$= \sum_{n=1}^{N}\left(\frac{1}{\sqrt{\det(\Sigma^{-1})}} \cdot \frac{1}{2\sqrt{\det(\Sigma^{-1})}} \Sigma \cdot \cancel{\det(\Sigma^{-1})}\right) - \frac{1}{2}\frac{tr[(x_n-\mu_{kn})^T \Sigma^{-1}(x_n-\mu_{kn})]}{d\Sigma^{-1}}$$

$$= \frac{N}{2}\Sigma - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu_{kn})(x_n-\mu_{kn})^T) = 0$$

$$\boxed{\hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{kn})(x_n-\mu_{kn})^T}$$

d) pooled diagonal covariance matrix.

$$p(x \mid \mu_k, \sigma^2) = \frac{1}{\prod_{d=1}^{D} \sqrt{2\pi \sigma_d^2}} \exp\left[-\frac{1}{2} \sum_{d=1}^{D} \left(\frac{x_d - \mu_{kd}}{\sigma_d}\right)^2\right]$$

$$\prod_{n=1}^{N} p(x_n \mid \mu_{k_n}, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\prod_{d=1}^{D} \sqrt{2\pi \sigma_d^2}} \exp\left[-\frac{1}{2} \sum_{d=1}^{D} \left(\frac{x_{nd} - \mu_{k_n,d}}{\sigma_d}\right)^2\right] =$$

$$\overset{\log}{\Longrightarrow} \sum_{n=1}^{N} \left(\log \frac{1}{\prod_{d=1}^{D} \sqrt{2\pi \sigma_d^2}} + \log\left(\exp\left[-\frac{1}{2} \sum_{d=1}^{D} \left(\frac{x_{nd} - \mu_{k_n d}}{\sigma_d}\right)^2\right]\right)\right) =$$

$$= \sum_{n=1}^{N} \left(-D \log \sqrt{2\pi} - \sum_{d=1}^{D} \log \sigma_d - \frac{1}{2} \sum_{d=1}^{D} \left(\frac{x_{nd} - \mu_{k_n,d}}{\sigma_d}\right)^2\right)$$

$\hat{\sigma}_d = \underset{\sigma_d}{\operatorname{argmax}} \{ L(\sigma_d) \}$ for all classes and dimension d

$$\frac{dL(\sigma_d)}{d\sigma_d} = \sum_{n=1}^{N} \left(-\frac{1}{\sigma_d} - \frac{1}{\cancel{2}} (x_{nd} - \mu_{k_n,d})^2 \cdot \left(-\frac{\cancel{2}}{\sigma_d^3}\right)\right) = 0 \,/\cdot \sigma_d^3$$

$$\sum_{n=1}^{N} \left(-\sigma_d^2 + (x_{nd} - \mu_{k_n,d})^2\right) = 0$$

$$-N\sigma_d^2 + \sum_{n=1}^{N} (x_{nd} - \mu_{k_n,d})^2 = 0$$

$$\boxed{\hat{\sigma}_d^2 = \frac{1}{N} \sum_{n=1}^{N} (x_{nd} - \mu_{k_n,d})^2}$$