1) (a)

1. Let l be likelihood

$$\log l = \sum_{i=1}^{m} \log\left(\frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\bar{x}_i-\bar{\mu})^T\Sigma^{-1}(\bar{x}_i-\bar{\mu})}\right)$$

where $\Sigma \Rightarrow$ Covariance Matrix, $n \Rightarrow$ Dimension of x_i $\forall i=1,2,\ldots,m$

$\mu \Rightarrow$ Mean vector

$$\Rightarrow \log l = -\sum_{i=1}^{m}\left(\log 2\pi^{n/2} + \frac{1}{2}\log|\Sigma| + \frac{1}{2}(\bar{x}_i-\bar{\mu})^T\Sigma^{-1}(\bar{x}_i-\bar{\mu})\right)$$

- To get maximize likelihood (log likelihood),

$$\boxed{\frac{\partial \log l}{\partial \bar{\mu}} = 0, \quad \frac{\partial \log l}{\partial \Sigma} = 0}$$

- $$\frac{\partial \log l}{\partial \bar{\mu}} = -\sum_{i=1}^{m}\left(\frac{1}{2|\Sigma|}\frac{\partial|\Sigma|}{\partial\bar{\mu}} + \frac{1}{2}\frac{\partial}{\partial\mu}\left((\bar{x}_i-\bar{\mu})^T\Sigma^{-1}(\bar{x}_i-\bar{\mu})\right)\right)$$

$$\Rightarrow \cancel{\frac{\partial \log l}{\partial \mu}}^{0} = -\sum_{i=1}^{m}\left(0 + \frac{1}{2}\left(-2\Sigma^{-1}(\bar{x}_i-\bar{\mu})\right)\right)$$

(As, $\frac{\partial}{\partial t}(t^TAt) = (A+A^T)t$, Here Σ^{-1} is A where Σ^{-1} is symmetric, hence derivative is $2\Sigma^{-1}(\bar{x}_i-\bar{\mu})$. Also we have $(\bar{x}_i-\bar{\mu})$ so extra minus comes) $\rightarrow$ (1)

$$\Rightarrow \boxed{\bar{\mu} = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i}$$

* $\dfrac{\partial \log l}{\partial \Sigma} = -\sum_{i=1}^{m}\left(\dfrac{1}{2}\Sigma^{-1} + \dfrac{1}{2}\dfrac{\partial}{\partial \Sigma}\left((\bar{x}_i-\bar{\mu})^T\Sigma^{-1}(\bar{x}_i-\bar{\mu})\right)\right)$

(As $\dfrac{\partial \log|A|}{\partial A} = (A^{-1})^T$) → (2)

$$\Rightarrow \frac{\partial \log l}{\partial \Sigma^{-1}} = -\sum_{i=1}^{m}\frac{1}{2}\Sigma + \frac{1}{2}(\bar{x}_i-\bar{\mu})(\bar{x}_i-\bar{\mu})^T$$

($\frac{\partial}{\partial \Sigma^{-1}}$ operator is used, as we have Σ^{-1} in RHS)

(As $(\bar{x}_i-\bar{\mu})^T\Sigma^{-1}(\bar{x}_i-\bar{\mu}) = tr((\bar{x}_i-\bar{\mu})(\bar{x}_i-\bar{\mu})^T\Sigma^{-1})$), ($tr(BCA) = tr(ABC)$)

$$\boxed{\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\bar{x}_i-\bar{\mu})(\bar{x}_i-\bar{\mu})^T}$$

$tr(x^TAx) = x^TAx$ as x^TAx is Scalar

(b)
MAP

* Lets assume Normal prior for $(\bar{\mu}_{MAP})_{Prior} \sim N(\mu, \Sigma)$ (Multivariate)

* Lets take $\mu = \mu_{MLE}$, $\Sigma = \Sigma_{MLE} \Rightarrow (\bar{\mu}_{MAP})_{Prior} \sim N(\mu_{MLE}, \Sigma_{MLE})$

* Let $\mu_{MLE} = \mu_0$, $\Sigma_{MLE} = \Sigma_0$

• $\text{argmax}\left(f\left(\dfrac{\theta}{x_1, x_2, \ldots, x_n}\right)\right) = \text{argmax}\left(f\left(\dfrac{x_1, x_2, \ldots, x_n}{\theta}\right) \cdot f(\theta)\right)$

μ: $\text{argmax}\left(N\left(\dfrac{\bar{\mu}}{\text{Data}}\right)\right) = \text{argmax}\left(N\left(\dfrac{\text{Data}}{\bar{\mu}}\right) \cdot N(\bar{\mu})\right)$

$\frac{27}{125}$

$$\Rightarrow \log\left(\operatorname{argmax}\left(N\left(\frac{\bar{\mu}}{\text{Data}}\right)\right)\right) = \sum_{i=1}^{m} \log\left(\frac{1}{2\pi^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\bar{x}_i-\bar{\mu})^T \Sigma^{-1}(\bar{x}_i-\bar{\mu})}\right) + \log\left(\frac{1}{2\pi^{D/2}|\Sigma_0|^{1/2}} e^{-\frac{1}{2}(\bar{\mu}-\bar{\mu}_0)^T \Sigma_0^{-1}(\bar{\mu}-\bar{\mu}_0)}\right)$$

$$\Rightarrow \frac{\partial \log\left(\operatorname{argmax}\left(N\left(\frac{\bar{\mu}}{\text{Data}}\right)\right)\right)}{\partial \bar{\mu}} = -\left(\sum_{i=1}^{m}\left(0 + \frac{1}{2|\Sigma|}\frac{\partial|\Sigma|}{\partial\bar{\mu}} + \frac{1}{2}\left(-2\Sigma^{-1}(\bar{x}_i-\bar{\mu}\right)\right) + \left(0 + \frac{1}{|\Sigma_0|^{1/2}}\frac{\partial|\Sigma_0|}{\partial\bar{\mu}} + \frac{1}{2}\left(-2\Sigma_0^{-1}(\bar{\mu}-\bar{\mu}\right)\right)\right)$$

Reason $\Rightarrow$ 1

$$\Rightarrow \frac{\partial \log(\,)}{\partial \bar{\mu}} \to 0 = -\sum_{i=1}^{m} -\Sigma^{-1}(\bar{x}_i-\bar{\mu}) + \Sigma_0^{-1}(\bar{\mu}-\bar{\mu}_0)$$

$$\Rightarrow \boxed{\bar{\mu}_{MAP} = \left(\Sigma_0^{-1} + m\Sigma^{-1}\right)\left(\Sigma_0^{(-)}\mu_0 + \sum_{i=1}^{m}\left(\Sigma^{-1}\bar{x}_i\right)\right)}$$

Q

2) $$E\left(\frac{1}{m}\sum_{i=1}^{m}(\bar{x}_i-\bar{\mu})(\bar{x}_i-\bar{\mu})^T\right)=\frac{1}{m}E\left(\sum_{i=1}^{m}(\bar{x}_i\bar{x}_i^T-\bar{\mu}\bar{x}_i-\bar{x}_i\bar{\mu}+\mu\mu^T)\right)$$

$$\Rightarrow\quad =\frac{1}{m}E\left(\sum_{i=1}^{m}(\bar{x}_i\bar{x}_i^T-\bar{\mu}\bar{\mu}^T)\right)$$

$$\Rightarrow\quad =\frac{1}{m}E\left(\sum(x_ix_i^T)-\cancel{m}\,m\left(\frac{1}{m}\sum\bar{x}_i\right)\left(\frac{1}{m}\sum x_i\right)^T\right)$$

$$\Rightarrow\quad =\frac{1}{m}\left(E\left(\sum(x_ix_i^T)\right)-\frac{1}{m}E\left(\sum\bar{x}_i\sum\bar{x}_i^T\right)\right)$$

$$\Rightarrow\quad =\frac{1}{m}\left(m\,\mu_{xx^T}-\frac{1}{m}\left(m\,\mu_{xx^T}+m(m-1)\mu_{x}\mu_{x^T}\right)\right)$$

$$\Rightarrow\quad =\frac{(m-1)}{m}\left(\mu_{xx^T}-\mu_x\mu_{x^T}\right)$$

$$\Rightarrow\quad \boxed{E(\Sigma)=\frac{(m-1)}{m}\sigma_{xx^T}}$$

* For Σ_x to be unbiased estimator, Σ has to be this

$$\boxed{\Sigma=\frac{1}{m-1}\sum_{i=1}^{m}(\bar{x}_i-\bar{\mu})(\bar{x}_i-\bar{\mu})^T}$$

30 Aug 2025, 11:09 pm