Given $\mathcal{X} = \{x^t\}_{t=1}^{N}$, $p(x^t|\theta) \sim N_d(\mu, \Sigma)$ (assumed)

$\forall\ x^t \sim N_d(\mu, \Sigma)$, $\theta = \{\mu, \Sigma\}$.

To find, estimators for $\mu\ (\hat{\mu})$ & $\Sigma\ (\hat{\Sigma})$, use MLE.

Now, $$l(\theta|x^t) = p(x^t|\theta)$$
$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(\frac{-1}{2}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right)$$
$$l(\theta|\mathcal{X}) = \prod_{t=1}^{N} l(\theta|x^t) \quad \text{(i.i.ds)}$$
$$= \prod_{t=1}^{N} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(\frac{-1}{2}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right)$$
$$\mathcal{L}(\theta|\mathcal{X}) = \log l(\theta|\mathcal{X}) = \sum_t \left[\frac{-d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

Now, for $\hat{\mu}$, find value of μ from $\frac{\partial \mathcal{L}}{\partial \mu} = 0$.

1) $$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_t \frac{\partial}{\partial \mu}\left[\frac{-d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$
$$= \sum_t \left[0 + 0 - \frac{1}{2} \times 2\Sigma^{-1}(x^t-\mu)\right] \qquad \left(\because \frac{\partial A^TBA}{\partial A} = 2BA\right)\begin{pmatrix} A=(x^t-\mu) \\ B=\Sigma^{-1} \end{pmatrix}$$
$$= \Sigma^{-1}\left[-\sum_t x^t + \sum_t \mu\right] = \Sigma^{-1}\left[N\mu - \sum_t x^t\right]$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = 0 \implies \Sigma^{-1}\left[N\mu - \sum_t x^t\right] = 0$$
$$\implies N\mu - \sum_t x^t = 0$$
$$\implies N\mu = \sum_t x^t \implies \boxed{\hat{\mu}_{MLE} = \frac{\sum_t x^t}{N}}$$

Now, for Σ, find value of Σ from $\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = 0$.

2) $$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \sum_t \frac{\partial}{\partial \Sigma}\left[\frac{-d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$
$$= \sum_t \left[\frac{-1}{2}\frac{\partial}{\partial \Sigma^{-1}}\log|\Sigma| - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}}\left[(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]\right]$$

Now, $\frac{\partial}{\partial A}\log(|A|) = (A^{-1})^T$, and $\frac{1}{|A|} = |A^{-1}|$

$$\implies = \frac{1}{2}\sum_t \left[\frac{\partial}{\partial \Sigma^{-1}}\log\left(\frac{1}{|\Sigma|}\right) - \frac{\partial}{\partial \Sigma^{-1}}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

$$\Rightarrow \quad = \frac{1}{2}\Sigma_t\left[\frac{\partial}{\partial \Sigma^{-1}} \log|\Sigma^{-1}| - \frac{\partial}{\partial \Sigma^{-1}}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

$$= \frac{1}{2}\Sigma_t\left[((\Sigma^{-1})^{-1})^T - \frac{\partial}{\partial \Sigma^{-1}}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

Now, $\Sigma^T = \Sigma$ (Σ is symmetric)

$$\Rightarrow \quad = \frac{1}{2}\sum_t\left[\Sigma + \frac{\partial}{\partial \Sigma^{-1}}(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right]$$

$$= \frac{1}{2}\sum_t\left[\Sigma - \frac{\partial}{\partial \Sigma^{-1}} tr\left((x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right)\right] \quad \text{— } (1)$$

($\because$ $(x^t-\mu)^T\Sigma^{-1}(x^t-\mu)$ is a scalar, &
If A is a 1×1 matrix / scalar, $A = tr(A)$,
where $tr(A)$ = trace = $\sum_i A_{i,i}$ (sum of diagonal values))

Now, $tr(ABC) = tr(CAB) = tr(BCA)$
(trace of product of matrices is invariant to cyclic permutations)

$$\Rightarrow \quad tr\big(\underbrace{(x^t-\mu)^T}_{A}\underbrace{(\Sigma^{-1})}_{B}\underbrace{(x^t-\mu)}_{C}\big) = tr\big(\underbrace{(x^t-\mu)}_{C}\underbrace{(x^t-\mu)^T}_{A}\underbrace{\Sigma^{-1}}_{B}\big)$$

$$= tr\big(\underbrace{\Sigma^{-1}}_{B}\underbrace{(x^t-\mu)}_{C}\underbrace{(x^t-\mu)^T}_{A}\big)$$

Now, $\frac{\partial}{\partial A} tr(AB) = B^T$. Assuming $A = \Sigma^{-1}$, $B = (x^t-\mu)(x^t-\mu)^T$,

$$\Rightarrow \frac{\partial}{\partial \Sigma^{-1}} tr\left((x^t-\mu)^T\Sigma^{-1}(x^t-\mu)\right) = \left[(x^t-\mu)(x^t-\mu)^T\right]^T$$

$$(1) \Rightarrow \frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \frac{1}{2}\sum_t\left[\Sigma - \left[(x^t-\mu)(x^t-\mu)^T\right]^T\right]$$

$$= \frac{1}{2}\sum_t\left[\Sigma - (x^t-\mu)^{T^T}(x^t-\mu)^T\right] \quad \left(\because (AB)^T = B^TA^T\right)$$

$$= \frac{1}{2}\sum_t\left[\Sigma - (x^t-\mu)(x^t-\mu)^T\right]$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = 0 \Rightarrow \Sigma_t(\Sigma) - \Sigma_t(x^t-\mu)(x^t-\mu)^T = 0$$

$$\Rightarrow N\Sigma = \Sigma_t(x^t-\mu)(x^t-\mu)^T$$

$$\Rightarrow \boxed{\hat{\Sigma}_{MLE} = \frac{\Sigma_t(x^t-\mu)(x^t-\mu)^T}{N}}$$