1. $f_K(x) = \frac{1}{\sqrt{(2\pi)^K |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$

Since Σ is positive definite. there exists invertible lower triangular matrix L s.t. $\Sigma = LL^T \Rightarrow \Sigma^{-1} = (LL^T)^{-1} = (L^T)^{-1} L^{-1}$ and $\det(\Sigma) = \det(L)\det(L^T)$

$= (\det(L))^2$

Let $y = L^{-1}(x-\mu) \Rightarrow x = \mu + Ly$

$\Rightarrow \det(L) = \sqrt{\det(\Sigma)} = |\Sigma|^{\frac{1}{2}}$

then $f_K(x) = \frac{1}{\sqrt{(2\pi)^K |\Sigma|}} \exp\left(-\frac{1}{2}(Ly)^T \Sigma^{-1}(Ly)\right) = \frac{1}{\sqrt{(2\pi)^K|\Sigma|}} \exp\left(-\frac{1}{2} y^T L^T (L^T)^{-1} L^{-1} Ly\right)$

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

$$\int_{\mathbb{R}^K} f_K(x)\, dx = \int_{\mathbb{R}^K} f_K(y)\, d(\mu + Ly) = \int_{\mathbb{R}^K} f_K(y) \cdot |\det(L)|\, dy$$

$$= \int_{\mathbb{R}^K} \frac{1}{\sqrt{(2\pi)^K |\Sigma|}} \exp\left(-\frac{1}{2} y^T y\right) |\Sigma|^{\frac{1}{2}}\, dy = \int_{\mathbb{R}^K} \frac{1}{(2\pi)^{\frac{K}{2}}} e^{-\frac{1}{2} y^T y}\, dy$$

$$= \int_{\mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}} \frac{1}{(2\pi)^{\frac{K}{2}}} e^{-\frac{1}{2}(y_1^2 + y_2^2 + \cdots + y_K^2)}\, dy$$

$$= \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2}\, dy_1\right)\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_2^2}\, dy_2\right) \cdots \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_K^2}\, dy_K\right)$$

$$= \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy\right)^K = 1^K = 1 \ \#$$

2. ① $\operatorname{trace}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} A_{ik} B_{ki}\right)$

$$\frac{\partial}{\partial A} \operatorname{trace}(AB) = \left[\frac{\partial}{\partial A_{ij}} \operatorname{trace}(AB)\right]_{i,j \in \{1,2,\dots,n\}}$$

$$= \left[\frac{\partial}{\partial A_{ij}} A_{ij} B_{ji}\right] = \left[B_{ji}\right] = B^T \ \#$$

② $x^T A x = \left[\sum_{i=1}^{n} x_i A_{i1}, \sum_{i=1}^{n} x_i A_{i2}, \dots, \sum_{i=1}^{n} x_i A_{in}\right] x = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{n} x_i A_{ij}\right) = \sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j A_{ij}$

$$\operatorname{trace}(xx^T A) = \operatorname{trace}\left(\begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2x_2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_nx_n \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}\right)$$

$$= \left(\sum_{i=1}^{n} x_1 x_i A_{i1}\right) + \left(\sum_{i=1}^{n} x_2 x_i A_{i2}\right) + \cdots + \left(\sum_{i=1}^{n} x_n x_i A_{in}\right)$$

$$= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} x_j x_i A_{ij}\right) = x^T A x$$

③ likelihood function $L(\mu,\Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)$

($x_1, x_2, \ldots, x_n \in n\times 1$ vector)

log likelihood function $\ell(\mu,\Sigma) = \ln L(\mu,\Sigma)$

$$= \sum_{i=1}^{n} \ln\left(\frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)\right)$$

$$= \sum_{i=1}^{n} -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= -\frac{kn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= -\frac{kn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\mathrm{trace}\left((x_i-\mu)(x_i-\mu)^T\Sigma^{-1}\right)$$

Claim: $\frac{\partial}{\partial y}(y^TAy) = (A+A^T)y$

$$y^TAy = \sum_{j=1}^{n}\sum_{i=1}^{n} y_i A_{ij} y_j$$

$$\frac{\partial}{\partial y_k}(y^TAy) = \sum_{i=1}^{n} y_i A_{ik} + \sum_{j=1}^{n} A_{kj} y_j = (y^TA)_k + (Ay)_k$$

$$= (A^Ty)_k + (Ay)_k = ((A+A^T)y)_k$$

Hence, $\frac{\partial}{\partial y}(y^TAy) = (A+A^T)y$

$= 2Ay$ if A is symmetric

Let $y_i = x_i - \mu$

$$\Rightarrow \frac{\partial}{\partial\mu}\ell(\mu,\Sigma) = -\frac{1}{2}\sum_{i=1}^{n}\left(-2\Sigma^{-1}(x_i-\mu)\right) = \sum_{i=1}^{n}\left(\Sigma^{-1}(x_i-\mu)\right) = \Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu) := 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\hat{\mu} = 0 \Rightarrow \underline{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i} \ \#$$

$$\frac{\partial}{\partial\Sigma}\ell(\mu,\Sigma) = \frac{\partial}{\partial\Sigma}\left(-\frac{kn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\mathrm{trace}\left((x_i-\mu)(x_i-\mu)^T\Sigma^{-1}\right)\right)$$

Let $\Lambda = \Sigma^{-1}$

$\Rightarrow |\Lambda| = \frac{1}{|\Sigma|}$

$$= \frac{\partial}{\partial\Lambda}\left(-\frac{kn}{2}\ln(2\pi) + \frac{n}{2}\ln|\Lambda^{-1}| - \frac{1}{2}\sum_{i=1}^{n}\mathrm{trace}\left((x_i-\mu)(x_i-\mu)^T\Lambda\right)\right)$$

$$= \frac{\partial}{\partial\Lambda}\left(\frac{n}{2}\ln|\Lambda^{-1}|\right) - \frac{\partial}{\partial\Lambda}\left(\frac{1}{2}\mathrm{trace}\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\Lambda\right)\right)$$

$$\overset{(*)}{=} \frac{n}{2}\cdot(\Lambda^{-1})^T - \frac{\partial}{\partial\Lambda}\left(\frac{1}{2}\mathrm{trace}\left(\Lambda\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)\right)$$

$$\overset{(**)}{=} \frac{n}{2}(\Lambda^{-1})^T - \frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)^T$$

$$= \frac{n}{2}(\Lambda^T)^{-1} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T(x_i-\mu) \quad := 0$$

$$\Rightarrow n(\Lambda^T)^{-1} = \sum_{i=1}^{n}(x_i-\mu)^T(x_i-\mu) \overset{(\,)^T}{\Rightarrow} n\Lambda^{-1} = \sum_{i=1}^{n}(x_i-\mu)^T(x_i-\mu)$$

$$\Rightarrow \Lambda^{-1} = \underline{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^T(x_i-\mu)} = \hat{\Sigma} \quad \#$$

$(*)$ $\frac{\partial}{\partial A}\ln|A| = (A^{-1})^T$

$|A| = \sum_{i,j}(-1)^{i+j}A_{ij}M_{ij}$ where M_{ij} is the (i,j) minor of A

$\frac{\partial}{\partial A_{ij}}|A| = (-1)^{i+j}M_{ij} = C_{ij}$, where C is the matrix of cofactors

By chain rule $\frac{\partial}{\partial A_{ij}}\ln|A| = \frac{1}{|A|}C_{ij} \Rightarrow \frac{\partial}{\partial A}\ln|A| = \frac{1}{|A|}C$

Also recall that $A^{-1} = \frac{1}{|A|}C^T \Rightarrow C^T = |A|A^{-1} \Rightarrow C = |A|(A^{-1})^T$

$\Rightarrow \frac{\partial}{\partial A}\ln|A| = \frac{1}{|A|}\left(|A|(A^{-1})^T\right) = (A^{-1})^T$

$(**)$ $\frac{\partial}{\partial A}\mathrm{trace}(AB) = B^T$ from above