1. Given

$$f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

where $x, \mu \in \mathbb{R}^k$, Σ is a k-by-k positive definite matrix and $|\Sigma|$ is its determinant.
Show that $\int_{\mathbb{R}^k} f(x)\,dx = 1$.

Let $y = x - \mu \in \mathbb{R}^k$, then $\int_{\mathbb{R}^k} f(x)\,dx = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} e^{-\frac{1}{2} y^T \Sigma^{-1} y}\,dy$

Since Σ is positive-definite, $\exists$ Cholesky decomposition $\Sigma = LL^T$, L is invertible.

Let $z = L^{-1} y \in \mathbb{R}^k \Rightarrow y = Lz \Rightarrow dy = |\det(L)|\,dz$.

Then $y^T \Sigma^{-1} y = (Lz)^T ((L^T)^{-1} L^{-1})(Lz) = z^T z = \|z\|^2$

Also $\because |\Sigma| = |\det L|^2 \quad \therefore |\det L| = |\Sigma|^{1/2}$

Therefore, $\int_{\mathbb{R}^k} e^{-\frac{1}{2} y^T \Sigma^{-1} y}\,dy = |\det(L)| \int_{\mathbb{R}^k} e^{-\frac{1}{2}\|z\|^2}\,dz = |\Sigma|^{1/2} \int_{\mathbb{R}^k} e^{-\frac{1}{2}\|z\|^2}\,dz$

$$= |\Sigma|^{1/2} \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{1}{2} z_i^2}\,dz_i$$

$$= |\Sigma|^{1/2} \prod_{i=1}^{k} \sqrt{2\pi}$$

$$= |\Sigma|^{1/2} (2\pi)^{k/2}$$

Thus, $\int_{\mathbb{R}^k} f(x)\,dx = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} e^{-\frac{1}{2} y^T \Sigma^{-1} y}\,dy = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} |\Sigma|^{1/2} (2\pi)^{k/2} = 1.$

2. Let A, B be n-by-n matrices and x be a n-by-1 vector.
(a) Show that $\frac{\partial}{\partial A} \mathrm{trace}(AB) = B^T$.
(b) Show that $x^T A x = \mathrm{trace}(x x^T A)$.
(b) Derive the maximum likelihood estimators for a multivariate Gaussian.

(a) $\mathrm{trace}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ji}$

$$= (A_{11} B_{11} + A_{12} B_{21} + \cdots + A_{1n} B_{n1}) + (A_{21} B_{12} + \cdots + A_{2n} B_{n2}) + \cdots + (A_{n1} B_{1n} + \cdots + A_{nn} B_{nn})$$

For each element A_{kl},

$$\frac{\partial}{\partial A_{kl}} \mathrm{trace}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial}{\partial A_{kl}} (A_{ij} B_{ji}) = B_{lk}$$

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

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

xx^T is a n-by-n matrix and $(xx^T)_{ij} = x_i x_j$.

$[(xx^T)A]_{ij} = \sum_{k=1}^{n} x_i x_k A_{jk}$

$$\text{trace}(xx^T A) = \sum_{k=1}^{n} x_1 x_k A_{1k} + \sum_{k=1}^{n} x_2 x_k A_{2k} + \cdots + \sum_{k=1}^{n} x_n x_k A_{nk}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j A_{ij}$$

$$= x^T A x$$

(c) Let $\exists$ N independent data point $x_1, \cdots, x_N \in \mathbb{R}^k$.

PDF of multivariate Gaussian : $p(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$

$$L(\mu, \Sigma) = \prod_{n=1}^{N} p(x_n \mid \mu, \Sigma)$$

$$\log L(\mu, \Sigma) = \sum_{n=1}^{N} \log p(x_n \mid \mu, \Sigma)$$

$$= \sum_{n=1}^{N} \left[-\frac{k}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu) \right]$$

$$= -\frac{Nk}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1}(x_n - \mu)$$

By (b), $(x_n - \mu)^T \Sigma^{-1}(x_n - \mu) = \text{trace}((x_n - \mu)(x_n - \mu)^T \Sigma^{-1})$

$$\frac{\partial}{\partial \mu} \log L(\mu, \Sigma) = -\frac{1}{2} \frac{\partial}{\partial \mu} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1}(x_n - \mu) = \frac{\partial}{\partial \mu} \sum_{n=1}^{N} \text{trace}((x_n - \mu)(x_n - \mu)^T \Sigma^{-1})$$

$$d[\text{trace}((x_n - \mu)(x_n - \mu)^T \Sigma^{-1})] = \text{trace}(d[(x_n - \mu)(x_n - \mu)^T] \Sigma^{-1})$$

$$= \text{trace}([-d\mu(x_n - \mu)^T - (x_n - \mu) d\mu^T] \Sigma^{-1})$$

$$= -\text{trace}(\Sigma^{-1}(x_n - \mu)^T d\mu) - \text{trace}(\Sigma^{-1}(x_n - \mu) d\mu^T)$$

$$= -2\, \text{trace}((\Sigma^{-1}(x_n - \mu))^T d\mu) = -2 \Sigma^{-1}(x_n - \mu) d\mu$$

$$\frac{\partial}{\partial \mu} \log L(\mu, \Sigma) = -\frac{1}{2} \sum_{n=1}^{N} \frac{\partial}{\partial \mu} \left(-2 \Sigma^{-1}(x_n - \mu) d\mu \right) = \sum_{n=1}^{N} \Sigma^{-1}(x_n - \mu) = 0$$

$$\Rightarrow \sum_{n=1}^{N} (x_n - \mu) = 0 \Rightarrow N\mu = \sum_{n=1}^{N} x_n \Rightarrow \hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\frac{\partial}{\partial \Sigma^{-1}} \log L(\mu, \Sigma) = \frac{\partial}{\partial \Sigma^{-1}} \left[-\frac{Nk}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu) \right]$$

$$= \frac{\partial}{\partial \Sigma^{-1}} \left(\frac{N}{2}\log|\Sigma| \right) - \frac{\partial}{\partial \Sigma^{-1}} \left(\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu) \right)$$

$$= \frac{N}{2}\left(\Sigma^{-1}\right)^{-1} - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T = 0$$

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

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n \quad , \quad \hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T$$