1. Show $\int_{\mathbb{R}^k} f(x)\, dx = 1$:

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

1) Let $y = \Sigma^{-\frac{1}{2}}(x-\mu) \implies x = \mu + \Sigma^{\frac{1}{2}} y$

   Jacobian : $\left|\frac{\partial x}{\partial y}\right| = \left|\Sigma^{\frac{1}{2}}\right| = |\Sigma|^{\frac{1}{2}}$

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

   $\qquad = \frac{1}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}\sum_i y_i^2}\, dy$

   $\implies \int_{\mathbb{R}^k} e^{-\frac{1}{2}\sum_i y_i^2}\, dy = \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{y_i^2}{2}}\, dy_i = (\sqrt{2\pi})^k = (2\pi)^{\frac{k}{2}}$

3) $\int_{\mathbb{R}^k} f(x)\, dx = \frac{(2\pi)^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}} = 1$

2. (a)
$$\frac{\partial}{\partial A}\, \text{trace}(AB) = B^T$$

   <Pf>

   $\text{trace}(AB) = \sum_i (AB)_{ii} = \sum_i \sum_j A_{ij} B_{ji}$

   $\dfrac{\partial}{\partial A_{pq}}\, \text{trace}(AB) = B_{qp}$

   Hence the gradient matrix is B^T. ※

(b)

$$x^T A x = \text{trace}(x \cdot x^T \cdot A)$$

$$x^T A x = \sum_{i,j} x_i A_{ij} x_j$$

Since A_{ij} is constant, $\quad x^T A x = \sum_{i,j} A_{ij}(x_j x_i)$

$$= \text{trace}(A x x^T)$$

By cyclic property of the trace, $x^T A x = \text{trace}(A x x^T)$

$$= \text{trace}(x x^T A). \; \text{\Large *}$$

(c)

    Let the data $x_1, \ldots, x_n$ be i.i.d samples from $N(\mu, \Sigma)$

$$L(\mu, \Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)}$$

$$l(\mu, \Sigma) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu).$$

1)

    Define sample scatter matrix:

$$S(\mu) = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$$

$$l(\mu, \Sigma) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\log(\Sigma^{-1} S(\mu))$$

$$\because \quad dS(\mu) = \sum_{i=1}^{n} \left[-d\mu(x_i - \mu)^T - (x_i - \mu)\,d\mu^T \right]$$

$$\therefore \quad dl = -\frac{1}{2}\text{trace}(\Sigma^{-1}\, dS(\mu))$$

$$= \frac{1}{2}\sum_{i=1}^{n}\text{trace}(\Sigma^{-1}\, d\mu(x_i - \mu)^T + \Sigma^{-1}(x_i - \mu)\,d\mu^T$$

$$= \sum_{i=1}^{n}(\Sigma^{-1}(x_i - \mu))^T\, d\mu$$

$$\frac{\partial l}{\partial \mu} = \Sigma^{-1}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i. \; \text{\Large *}$$

2)

With $\mu = \hat{\mu}$

$$\ell(\Sigma) = -\frac{n}{2} \log|\Sigma| - \frac{1}{2}(\Sigma^{-1}S), \quad S = \sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T$$

Since $d\log|\Sigma| = \text{trace}(\Sigma^{-1}d\Sigma), \quad d(\Sigma^{-1}) = -\Sigma^{-1}(d\Sigma)\Sigma^{-1}$

$$d\ell = -\frac{n}{2}\text{trace}(\Sigma^{-1}d\Sigma) - \frac{1}{2}\text{trace}(d(\Sigma^{-1})S)$$

$$= -\frac{n}{2}\text{trace}(\Sigma^{-1}d\Sigma) + \frac{1}{2}\text{trace}(\Sigma^{-1}d\Sigma\Sigma^{-1}S)$$

$$= \frac{1}{2}\text{trace}((-n\Sigma^{-1} + \Sigma^{-1}S\Sigma^{-1})d\Sigma)$$

For any symmetric $d\Sigma$, set $d\ell = 0$

$$-n\Sigma^{-1} + \Sigma^{-1}S\Sigma^{-1} = 0 \Rightarrow S = n\Sigma$$

Thus $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T$ ✳