1. Since Σ is positive define, there exists a matrix A such that $\Sigma = AA^T$

Let $y = A^{-1}(x - \mu)$

Then $(x-\mu)^T \Sigma^{-1} (x-\mu) = (Ay)^T (AA^T)^{-1} (Ay) = y^T y = \|y\|^2$

$dx = |\det A|\, dy$

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

By Gaussian integral, $\int_{\mathbb{R}} e^{-\frac{1}{2}\|y\|^2} dy = \sqrt{(2\pi)^k}$

$\Rightarrow \int_{\mathbb{R}^k} f(x) dx = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \times \sqrt{(2\pi)^k} \times |\Sigma|^{\frac{1}{2}}$

$= 1$

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

$\frac{\partial}{\partial A_{xy}} tr(AB) = \frac{\partial}{\partial A_{xy}} \sum_{ij} A_{ij} B_{ji} = \frac{\partial}{\partial A_{xy}} A_{xy} B_{yx} = B_{yx}$

So $\frac{\partial}{\partial A} tr(AB) = B^T$

(b) $x^T A x \in \mathbb{R}$

$= tr(x^T A x)$

Since $tr(AB) = tr(BA)$, for all $A \in \mathbb{R}^{m \times n}$ $B \in \mathbb{R}^{n \times m}$

$tr(x^T A x) = tr(x x^T A)$

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

$\theta = (\mu, \Sigma)$

$$\theta^* = \arg\max_\theta L(\theta) = \arg\max_\theta \ln L(\theta)$$

$$\ln L(\theta) = -\frac{N}{2} \ln\left((2\pi)^k |\Sigma|\right) - \frac{1}{2} \sum_{i=1}^{N} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu) = l(\mu, \Sigma)$$

By (a)、(b) $\frac{\partial l}{\partial \mu} = \frac{\partial}{\partial \mu}\left(-\frac{1}{2} \sum_i (x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right) = -\frac{1}{2}\frac{\partial}{\partial \mu} tr\left((x_i-\mu)(x_i-\mu)^T \Sigma^{-1}\right)$

$$= \frac{1}{2} \Sigma \sum_{i=1}^{N} (x_i-\mu) = 0$$

$$\Rightarrow \mu^* = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\frac{\partial l}{\partial \Sigma} = -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \sum_{i=1}^{N} (x_i-\mu)(x_i-\mu)^T \Sigma^{-1} = 0$$

$$\Rightarrow \Sigma^* = \frac{1}{N} \sum_{i=1}^{N} (x_i-\mu)(x_i-\mu)^T$$

3、What's the use of cross-entropy loss, compare with MSE.