1. Since Σ is a $k \times k$ positive define matrix

We can let $\Sigma = LL^T$ where L is a lower triangular matrix

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

We have $dx = |\det(L)|\, dy = |\Sigma|^{\frac{1}{2}} dy$. (Since $\det(LL^T) = \det(L)^2 = |\Sigma|$)

And $(x-\mu)^T \Sigma^{-1} (x-\mu) = y^T L^T (LL^T)^{-1} L y = y^T y$

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

$$= \frac{1}{\sqrt{(2\pi)^k}} \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{y_i^2}{2}} dy_i = \frac{1}{\sqrt{(2\pi)^k}} \cdot (\sqrt{2\pi})^k = 1 \quad \#$$

2. (a) By def. $\operatorname{trace}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ji}$

Therefore, $\frac{\partial}{\partial A_{xy}} \operatorname{trace}(AB) = B_{yx} \Rightarrow \frac{\partial}{\partial A} \operatorname{trace}(AB) = B^T$

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

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

Therefore, $x^T A x = \operatorname{trace}(x x^T A)$

(c) Given: $x_1, \dots, x_n \sim N(\mu, \Sigma)$

Log-Likelihood: $\ln L(\mu, \Sigma) = \frac{-nk}{2} \ln(2\pi) - \frac{n}{2} \ln|\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$

For μ: $\frac{\partial}{\partial \mu} \ln L = \Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$

For Σ: use $d(\ln|\Sigma|) = \operatorname{trace}(\Sigma^{-1} d\Sigma)$

$d(\operatorname{trace}(\Sigma^{-1} S \Sigma)) = -\operatorname{trace}(\Sigma^{-1}(d\Sigma)\Sigma^{-1} S)$ where $S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^T$

$\frac{-n}{2} \operatorname{trace}(\Sigma^{-1} d\Sigma) + \frac{n}{2} \operatorname{trace}(\Sigma^{-1} S \Sigma^{-1} d\Sigma) = 0$

$\Sigma^{-1} = \Sigma^{-1} S \Sigma^{-1} \Rightarrow \hat{\Sigma} = S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$

3. 我上次作業有用到Bootstrap Aggregating，其中使用隨機子空間法，令原始空間維度為p，則子空間維度通常為$\lfloor \sqrt{p} \rfloor$， 我不太明白這樣選擇子空間維度的用意，有什麼特性或好處嗎？