(discuss with roommate).

1. Given $f(x) = \frac{1}{((2\pi)^k |\Sigma|)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$, where $x, \mu \in \mathbb{R}^k$,

Σ is a k-by-k positive define matrix and $|\Sigma|$ is its determinant.

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

Proof.

Since the hypothesis given, we can rewrite $\int_{\mathbb{R}^k} f(x)\, dx$

as $\int_{\mathbb{R}^k} ((2\pi)^k |\Sigma|)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx$.

Let $u = \sigma^{-\frac{1}{2}}(x-\mu)$. we get $(x-\mu)^T \Sigma^{-1}(x-\mu) = u^T u$ and $dx = \sqrt{|\Sigma|}\, du$.

Then we have
$$\int_{\mathbb{R}^k} ((2\pi)^k |\Sigma|)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx = (2\pi)^{-\frac{k}{2}} \cdot \int_{\mathbb{R}^k} e^{-\frac{u^T u}{2}} du$$
$$= (2\pi)^{-\frac{k}{2}} \cdot \int_{\mathbb{R}^k} e^{-\frac{1}{2}(\sum_{i=1}^k u_i^2)} du$$
$$= (2\pi)^{-\frac{k}{2}} \cdot \prod_{i=1}^k \int_{\mathbb{R}} e^{-\frac{1}{2} \cdot u_i^2} du_i$$
$$= (2\pi)^{-\frac{k}{2}} \cdot \prod_{i=1}^k \sqrt{2\pi} = (2\pi)^{-\frac{k}{2}} \cdot (2\pi)^{\frac{k}{2}} = 1,$$

where u_i, for $i = 1, 2, \cdots, k$ are components of u.

2. Let A, B be n-by-n matrices and x be a n-by-1 vector.

(a) Show that $\frac{\partial}{\partial A} \operatorname{trace}(AB) = B^T$.

Since
$$\operatorname{trace}(AB) = \sum_{i=1}^n (A, B)_{i,i}$$
$$= \sum_{i=1}^n \sum_{j=1}^n A_{i,j} B_{j,i},$$

we have
$$\frac{\partial}{\partial A_{k,l}} \operatorname{trace}(AB) = \frac{\partial}{\partial A_{k,l}} \sum_{i=1}^n \sum_{j=1}^n A_{i,j} B_{j,i}$$
$$= B_{l,k}.$$

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

(b) Show that $x^T A x = \text{trace}(x x^T A)$.

We prove this by expanding both sides.

(Left)

$$x^T A x = \sum_{i=1}^{n} x_i (Ax)_i$$

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

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

(Right)

Since $\text{trace}(M_1 M_2) = \text{trace}(M_2 M_1)$ for any M_1, M_2 are n-by-n matrices, for all $n \in \mathbb{N}$.

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

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

Therefore, the left side is equal to right side.

(c) discuss with ChatGPT and roommate.

Derive the maximum likelihood estimators for a multivariate Gaussian.

Given $x_1, x_2, \cdots, x_m$ are i.i.d. from $N(\mu, \Sigma)$ where $\mu \in \mathbb{R}^n$ the mean vector

$\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix.

The log-likelihood function

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

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

Setting to zero.

we have $\sum_{i=1}^{m} (x_i - \mu) = 0$ such that $\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} x_i$.

By using matrix Identities,

we have $\frac{\partial}{\partial \Sigma} \log |\Sigma| = (\Sigma^{-1})^T = \Sigma^{-1}$

$$\frac{\partial}{\partial \Sigma} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) = -\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T \Sigma^{-1}.$$

Then $\frac{\partial l}{\partial \Sigma} = -\frac{m}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left(\sum_{i=1}^{m} (x_i - \mu)(x_i - \mu)^T \right) \Sigma^{-1}$

Setting to zero, we get $-m \Sigma^{-1} + \Sigma^{-1} \left(\sum_{i=1}^{m} (x_i - \mu)(x_i - \mu)^T \right) \Sigma^{-1} = 0$.

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

$$\hat{\Sigma} = \frac{1}{m} \sum_{i=1}^{m} (x_i - \hat{\mu})(x_i - \hat{\mu}).$$