1.

Theorem

Let $\mu \in \mathbb{R}^k$ and let $\Sigma \in \mathbb{R}^{k\times k}$ be a symmetric positive definite matrix.

Define

$$f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^k$$

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

Lemma 1

$$I = \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\,dt = \sqrt{2\pi}$$

pf of lemma 1:

Since the integrand is nonnegative,

by Tonelli's thm $\Rightarrow$ $I^2 = \int_{\mathbb{R}^2} e^{-(t^2+s^2)/2}\,dt\,ds$

switch to polar coordinates $(r,\theta) \Rightarrow$ $= \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} \cdot r\,dr\,d\theta = 2\pi \int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr$

let $u = \frac{r^2}{2} \Rightarrow du = r\,dr$ then $I^2 = 2\pi \int_0^{\infty} e^{-u}\,du = 2\pi$

$\Rightarrow I = \sqrt{2\pi}$

*

Lemma 2:

$$\int_{\mathbb{R}^k} e^{-\frac{\|y\|^2}{2}}\,dy = (2\pi)^{\frac{k}{2}}, \quad \text{where } y \in \mathbb{R}^k$$

pf of Lemma 2:

Since $e^{-\frac{\|y\|^2}{2}} = \prod_{i=1}^{k} e^{-\frac{y_i^2}{2}}$, by Fubini's thm,

$$\int_{\mathbb{R}^k} e^{-\frac{\|y\|^2}{2}}\,dy = \prod_{i=1}^{k} \int_{\mathbb{R}} e^{-\frac{y_i^2}{2}}\,dy_i \underset{\text{by Lemma 1}}{=} (\sqrt{2\pi})^k = (2\pi)^{\frac{k}{2}}$$

*

pf of Theuem :

Since Σ is symmetric and positive definite, $\exists\ \Sigma^{\frac{1}{2}}$

s.t. $\Sigma = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}$, $\Sigma^{-1} = (\Sigma^{\frac{1}{2}})^{-1}(\Sigma^{\frac{1}{2}})^{-1}$, $|\Sigma^{\frac{1}{2}}| = \sqrt{|\Sigma|}$

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

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

Compute the quadratic fam:

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = (\Sigma^{\frac{1}{2}}y)^T\Sigma^{-1}(\Sigma^{\frac{1}{2}}y) = y^Ty = \|y\|^2$$

substitute $x = \mu + \Sigma^{\frac{1}{2}}y$ and $dx = \sqrt{|\Sigma|}\,dy$ into the integral

$$\Rightarrow \int_{\mathbb{R}^k} f(x)\,dx = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$$

$$= \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}\|y\|^2\right)\sqrt{|\Sigma|}\,dy$$

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

by Lemma 2, $\int_{\mathbb{R}^k} e^{-\frac{\|y\|^2}{2}}\,dy = (2\pi)^{\frac{k}{2}}$

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

#

2.

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

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

pf:

Let

$$f(A) = tr(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ji}$$

then $\frac{\partial f}{\partial A_{ij}} = B_{ji}$

by def. of the matrix derivative,

$$\frac{\partial f}{\partial A} = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \cdots & \frac{\partial f}{\partial A_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{n1}} & \cdots & \frac{\partial f}{\partial A_{nn}} \end{bmatrix} = B^T$$

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

(b) Show that $x^TAx = \text{trace}(xx^TA)$

pf: $\because$ x^TAx is a scalar $\Rightarrow$ $tr(x^TAx) = x^TAx$

$\because$ $tr(x^TAx) = tr(Axx^T) = tr(xx^TA)$ (use the cyclic property of trace)

hence, $x^TAx = tr(xx^TA)$ ※

(c) Derive the maximum likelihood estimators for a multivariate Gaussian

let independent samples $x_i \sim N(\mu, \Sigma)$, $x_1, \dots, x_N \in \mathbb{R}^k$

and Σ be a positive definite matrix.

The joint pdf is:

$$L(\mu, \Sigma) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right]$$

The log-likelihood is

$$\ell(\mu, \Sigma) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln(\Sigma) - \frac{1}{2} \sum_{i=1}^{N} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu)$$

by (b), $(x_i-\mu)^T \Sigma^{-1} (x_i-\mu) = tr\left(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\right)$

hence $$\ell(\mu, \Sigma) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln(\Sigma) - \frac{1}{2} tr\left(\Sigma^{-1} \sum_{i=1}^{N} (x_i-\mu)(x_i-\mu)^T\right)$$

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

Then
$$\ell(\mu, \Sigma) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln|\Sigma| - \frac{1}{2} tr\left(\Sigma^{-1} S(\mu)\right)$$

We have $$\frac{\partial}{\partial \mu} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu) = -2\Sigma^{-1}(x_i-\mu)$$

$$\Rightarrow \nabla_\mu \ell(\mu, \Sigma) = -\frac{1}{2}(-2) \sum_{i=1}^{N} \Sigma^{-1}(x_i-\mu) = \Sigma^{-1} \sum_{i=1}^{N} (x_i-\mu)$$

Set $\sum_{i=1}^{N} (x_i-\hat{\mu}) = 0 \Rightarrow \underline{\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i}$ ※

Let $S = S(\hat{u})$

note that $\frac{\partial \ln|\Sigma|}{\partial \Sigma} = (\Sigma^{-1})^T = \Sigma^{-1}$, $\frac{\partial\, tr(\Sigma^{-1}S)}{\partial \Sigma} = -(\Sigma^{-1}S\Sigma^{-1})^T = -\Sigma^{-1}S\Sigma^{-1}$

then

$$\nabla_\Sigma \ell(\hat{u}, \Sigma) = -\frac{N}{2}\Sigma^{-1} - \frac{1}{2}(-\Sigma^{-1}S\Sigma^{-1}) = -\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}S\Sigma^{-1}$$

Setting $-\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}S\Sigma^{-1} = 0$

$$\Rightarrow \Sigma^{-1}S = NI_k \Rightarrow \underline{\hat{\Sigma} = \frac{1}{N}S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{u})(x_i - \hat{u})^T}$$

#