

MACHINE LEARNING

ASSIGNMENT 5

CHANG Yung-Hsuan (張永璿)

111652004

eiken.sc11@nycu.edu.tw

October 1, 2025

1. Given

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

where $x, \mu \in \mathbb{R}^k$ and $\Sigma \in \mathcal{M}_{k \times k}$ is positive definite. Show that

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

Proof. Let $u = \sigma^{-\frac{1}{2}}(x - \mu)$. Then,

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

Thus, the integral becomes

$$\begin{aligned} \int_{\mathbb{R}^k} f(x) \, dx &= \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \cdot \exp\left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}\right) dx \\ &= \frac{1}{\sqrt{(2\pi)^k}} \cdot \int_{\mathbb{R}^k} \exp\left(-\frac{u^T u}{2}\right) du \\ &= \frac{1}{\sqrt{(2\pi)^k}} \cdot \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2} \cdot \sum_{i=1}^k u_i^2\right) du \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{(2\pi)^k}} \cdot \prod_{i=1}^k \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \cdot u_i^2\right) du_i \\
&= \frac{1}{\sqrt{(2\pi)^k}} \cdot \prod_{i=1}^k \sqrt{2\pi} \\
&= 1,
\end{aligned}$$

where u_i 's are components of u , i.e.,

$$u = (u_1, u_2, \dots, u_k).$$

2. Let $A, B \in \mathcal{M}_{n \times n}$ and let x be a vector with n components.

a. Show that

$$\frac{\partial}{\partial A} \operatorname{tr}(AB) = B^T. \quad (1)$$

Proof. We have

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{i=1}^n (AB)_{i,i} \\
&= \sum_{i=1}^n \sum_{j=1}^n A_{i,j} B_{j,i}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial}{\partial A_{p,q}} \operatorname{tr}(AB) &= \frac{\partial}{\partial A_{p,q}} \sum_{i=1}^n \sum_{j=1}^n A_{i,j} B_{j,i} \\
&= B_{q,p}.
\end{aligned}$$

Therefore,

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

b. Show that

$$x^T A x = \operatorname{tr}(x x^T A). \quad (2)$$

Proof. We firstly expand the left-hand side of (2),

$$\begin{aligned}
x^T A x &= \sum_{i=1}^n x_i (A x)_i \\
&= \sum_{i=1}^n x_i \sum_{j=1}^n A_{i,j} x_j \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i A_{i,j} x_j.
\end{aligned}$$

For the right-hand side of (2), we know that

$$\text{tr}(M_1 M_2) = \text{tr}(M_2 M_1)$$

for any square matrices with the same dimension M_1 and M_2 . Thus,

$$\begin{aligned}
\text{tr}(x x^T A) &= \text{tr}(A x x^T) \\
&= \sum_{i=1}^n \sum_{j=1}^n A_{i,j} (x x^T)_{j,i} \\
&= \sum_{i=1}^n \sum_{j=1}^n A_{i,j} x_i x_j.
\end{aligned}$$

As the left-hand side equals the right-hand side, this proves (2).

- c. Derive the maximum likelihood estimators for a multivariate Gaussian random variable.

Solution. Let $X_1, \dots, X_n \sim \mathcal{N}(\mu, \Sigma)$. The log-likelihood function is

$$\begin{aligned}
\ell(\mu, \Sigma) &= -\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \\
&= -\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr}((x_i - \mu)(x_i - \mu)^T \Sigma^{-1}) \\
&= -\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr}((x_i - \mu)(x_i - \mu)^T \Sigma^{-1}) \\
&= -\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T) \tag{3}
\end{aligned}$$

We first look for $\hat{\mu}$. We have

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu}(\mu, \Sigma) &= \frac{\partial}{\partial \mu} \left(-\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma^{-1}(x_i - \mu)(x_i - \mu)^T) \right) \\
&= -\frac{1}{2} \cdot \sum_{i=1}^n \frac{\partial}{\partial \mu} \left(\text{tr}(\Sigma^{-1}(x_i - \mu)(x_i - \mu)^T) \right) \\
&= -\frac{1}{2} \cdot \sum_{i=1}^n \frac{\partial}{\partial M} (\text{tr}(\Sigma^{-1}M)) \cdot \frac{\partial M}{\partial \mu} \\
&= -\frac{1}{2} \sum_{i=1}^n \left((\Sigma^{-1})^T \right) \cdot (-2(x_i - \mu)) \\
&= \sum_{i=1}^n \left((\Sigma^{-1})^T \right) \cdot (x_i - \mu) \\
&= \sum_{i=1}^n \Sigma^{-1} \cdot (x_i - \mu), \tag{4}
\end{aligned}$$

where $M = (x_i - \mu)(x_i - \mu)^T$ for simplicity, the fourth equation comes from (1), and the last equation comes from the symmetry of Σ . Set (4) to 0, we have

$$\sum_{i=1}^n \Sigma^{-1} \cdot (x_i - \mu) = 0,$$

which implies

$$\sum_{i=1}^n (x_i - \mu) = 0,$$

and thus

$$\begin{aligned}
\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i \\
&= \bar{x} \tag{5}
\end{aligned}$$

Now, we plug (5) in (3) to look for $\hat{\Sigma}$. We have

$$\begin{aligned}
\frac{\partial \ell}{\partial \Sigma}(\hat{\mu}, \Sigma) &= \frac{\partial}{\partial \Sigma} \left(-\frac{nk}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \right) \right) \\
&= \frac{\partial}{\partial \Sigma} \left(-\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \right) \right) \\
&= -\frac{n}{2} \cdot \frac{1}{\Sigma^T} - \frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial \Sigma} \left(\text{tr} \left(\Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \right) \right) \\
&= -\frac{n}{2} \cdot \Sigma^{-1} + \frac{1}{2} \sum_{i=1}^n \Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \Sigma^{-1}, \tag{6}
\end{aligned}$$

where the last equation comes from the symmetry of Σ , and see (7) for the partial derivative identity. Set (6) to 0, we have

$$-\frac{n}{2} \cdot \Sigma^{-1} + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \Sigma^{-2} = 0,$$

which implies

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

and thus

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

d. Show that the equation

$$\frac{\partial}{\partial \Sigma} \left(\text{tr} \left(\Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \right) \right) = -\Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})^T \Sigma^{-1} \tag{7}$$

in (6) holds.

Proof. Let $u = x_i - \bar{x}$. Then,

$$\text{tr}(\Sigma^{-1} u u^T) = \sum_{i=1}^n \sum_{j=1}^n (\Sigma^{-1})_{i,j} (u u^T)_{j,i}.$$

Thus,

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_{p,q}} \text{tr}(\Sigma^{-1}uu^T) &= \frac{\partial}{\partial \Sigma_{p,q}} \sum_{i=1}^n \sum_{j=1}^n (\Sigma^{-1})_{i,j} (uu^T)_{j,i} \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial}{\partial (\Sigma^{-1})_{i,j}} (\Sigma^{-1})_{i,j} (uu^T)_{j,i} \cdot \frac{\partial (\Sigma^{-1})_{i,j}}{\partial \Sigma_{p,q}} \\
&= \sum_{i=1}^n \sum_{j=1}^n (uu^T)_{j,i} \cdot \left(-(\Sigma^{-1})_{i,p} (\Sigma^{-1})_{q,j} \right) \\
&= - \sum_{i=1}^n \sum_{j=1}^n (\Sigma^{-1})_{q,j} (uu^T)_{j,i} (\Sigma^{-1})_{i,p} \\
&= -(\Sigma^{-1}uu^T\Sigma^{-1})_{q,p} \\
&= -(\Sigma^{-1}uu^T\Sigma^{-1})_{p,q}
\end{aligned} \tag{8}$$

where the third equation follows from (9) and the last comes from the symmetry of $\Sigma^{-1}uu^T\Sigma^{-1}$.

e. Show that the equation

$$\frac{\partial (\Sigma^{-1})_{i,j}}{\partial \Sigma_{p,q}} = -(\Sigma^{-1})_{i,p} (\Sigma^{-1})_{q,j} \tag{9}$$

in (8) holds.

Proof. We first start from the definition of inverse

$$\sum_{\ell=1}^n \Sigma_{a,\ell} (\Sigma^{-1})_{\ell,b} = \delta_{a,b}.$$

Then, we differentiate both sides with respect to $\Sigma_{p,q}$ to obtain

$$\begin{aligned}
&\frac{\partial}{\partial \Sigma_{p,q}} \sum_{\ell=1}^n \Sigma_{a,\ell} (\Sigma^{-1})_{\ell,b} = \frac{\partial}{\partial \Sigma_{p,q}} \delta_{a,b} \\
\Rightarrow &\sum_{\ell=1}^n \frac{\partial \Sigma_{a,\ell}}{\partial \Sigma_{p,q}} (\Sigma^{-1})_{\ell,b} + \sum_{\ell=1}^n \Sigma_{a,\ell} \frac{\partial (\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0 \\
\Rightarrow &\sum_{\ell=1}^n \delta_{a,p} \delta_{\ell,q} (\Sigma^{-1})_{\ell,b} + \sum_{\ell=1}^n \Sigma_{a,\ell} \frac{\partial (\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0.
\end{aligned}$$

Thus,

$$\delta_{a,p}(\Sigma^{-1})_{q,b} + \sum_{\ell=1}^n \Sigma_{a,\ell} \frac{\partial(\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0.$$

Then, we multiply both sides on the left by $(\Sigma^{-1})_{i,a}$ and sum over a , we have

$$\sum_{a=1}^n (\Sigma^{-1})_{i,a} \delta_{a,p}(\Sigma^{-1})_{q,b} + \sum_{a=1}^n (\Sigma^{-1})_{i,a} \sum_{\ell=1}^n \Sigma_{a,\ell} \frac{\partial(\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0$$

$$(\Sigma^{-1})_{i,p}(\Sigma^{-1})_{q,b} + \sum_{a=1}^n (\Sigma^{-1})_{i,a} \sum_{\ell=1}^n \Sigma_{a,\ell} \frac{\partial(\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0$$

$$(\Sigma^{-1})_{i,p}(\Sigma^{-1})_{q,b} + \sum_{\ell=1}^n \sum_{a=1}^n (\Sigma^{-1})_{i,a} \Sigma_{a,\ell} \frac{\partial(\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0$$

$$(\Sigma^{-1})_{i,p}(\Sigma^{-1})_{q,b} + \sum_{\ell=1}^n \delta_{i,\ell} \frac{\partial(\Sigma^{-1})_{\ell,b}}{\partial \Sigma_{p,q}} = 0$$

$$(\Sigma^{-1})_{i,p}(\Sigma^{-1})_{q,b} + \frac{\partial(\Sigma^{-1})_{i,b}}{\partial \Sigma_{p,q}} = 0.$$

3. There are unanswered questions from the lecture, and there are likely more questions we haven't covered. Take a moment to think about these questions. Write down the ones you find important, confusing, or interesting.

Answer. How do we calculate the estimators for those $\Sigma_1 \neq \Sigma_2$? Do we really calculate them or we just code so that we see the results in practice?