

1. $\therefore \Sigma$ is positive definite $\Rightarrow \Sigma = R^T R$ for some $R \in \mathbb{R}^{k \times k}$ and R is also positive (R^T is also positive)

$$\Rightarrow R^T = S^{-1} \Lambda S, \text{ where } \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_k \end{pmatrix}, \lambda_1, \dots, \lambda_k \text{ are eigenvalues of } R^T$$

$$\Rightarrow R^{-1} = \sum_{n=1}^k S^{-1} \Lambda_n S, \text{ where } \Lambda_n = \begin{cases} \Lambda_{nn} = \lambda_n \\ 0, \text{ otherwise} \end{cases}$$

$$\begin{aligned} \therefore (x-\mu)^T \Sigma^{-1} (x-\mu) &= (R^{-1}(x-\mu))^T R^{-1}(x-\mu) = \|R^{-1}(x-\mu)\|^2 \\ &= \left\| \sum_{n=1}^k S^{-1} \Lambda_n S (x-\mu) \right\|^2 \\ &= \sum_{n=1}^k \|S^{-1} \Lambda_n S (x-\mu)\|^2 \end{aligned}$$

$$\begin{aligned} \Rightarrow \int_{\mathbb{R}^k} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} dx &= \int_{\mathbb{R}^k} e^{-\frac{1}{2} \sum_{n=1}^k \|S^{-1} \Lambda_n S (x-\mu)\|^2} dx \\ &= \int_{\mathbb{R}^k} e^{-\frac{1}{2} \|S^{-1} \Lambda_1 S (x-\mu)\|^2} \cdot e^{-\frac{1}{2} \|S^{-1} \Lambda_2 S (x-\mu)\|^2} \cdot \dots \cdot e^{-\frac{1}{2} \|S^{-1} \Lambda_k S (x-\mu)\|^2} dx \\ &= \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2} \lambda_1 (x-\mu_1)^2} dx_1 \right) \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2} \lambda_2 (x-\mu_2)^2} dx_2 \right) \cdot \dots \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2} \lambda_k (x-\mu_k)^2} dx_k \right) \\ &= \left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_1 (x-\mu_1)^2}{\sqrt{2}}\right)^2} dx_1 \right) \left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_2 (x-\mu_2)^2}{\sqrt{2}}\right)^2} dx_2 \right) \dots \left(\int_{-\infty}^{\infty} e^{-\left(\frac{\lambda_k (x-\mu_k)^2}{\sqrt{2}}\right)^2} dx_k \right) \\ &= \frac{\sqrt{(2\pi)^k}}{\lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_k} \quad \text{--- (1)} \end{aligned}$$

$\therefore \Sigma = R^T R \Rightarrow \frac{1}{\lambda_1^2}, \frac{1}{\lambda_2^2}, \dots, \frac{1}{\lambda_k^2}$ are the eigenvalues of Σ

$$\Rightarrow |\Sigma| = \frac{1}{(\lambda_1 \lambda_2 \dots \lambda_k)^2} \quad \text{--- (2)}$$

$$\therefore \int_{\mathbb{R}^k} f(x) dx = \int \frac{(\lambda_1 \lambda_2 \dots \lambda_k)^2}{(2\pi)^k} \cdot \frac{\sqrt{(2\pi)^k}}{\lambda_1 \lambda_2 \dots \lambda_k} = 1 \quad \text{by (1) and (2)}$$

2. (a)

$$\text{Let } A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & & \vdots \\ \vdots & & \ddots & \\ b_{n1} & \dots & \dots & b_{nn} \end{pmatrix}$$

$$\text{trace}(AB) = \sum_{k=1}^n \left(\sum_{s=1}^n a_{ks} b_{sk} \right)$$

$$\Rightarrow \frac{\partial}{\partial a_{ij}} \text{trace}(AB) = b_{ji}, \text{ for } 1 \leq i, j \leq n$$

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

(b)

$$X^T A X = (x_1 \ x_2 \ \dots \ x_n) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{1 \leq i, j \leq n} a_{ij} x_i x_j$$

$$X X^T A = \begin{pmatrix} x_1 x_1 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2 x_2 & & \vdots \\ \vdots & & \ddots & \\ x_n x_1 & \dots & \dots & x_n x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix}$$

$$\therefore \text{trace}(X X^T A) = \sum_{k=1}^n \sum_{s=1}^n x_k x_s a_{sk} = \sum_{1 \leq i, j \leq n} a_{ij} x_i x_j$$

$$\Rightarrow X^T A X = \text{trace}(X X^T A)$$

(c)

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

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

$$\text{Let } \ell(\theta) = \ln L(\theta) = \frac{N}{2} \ln \frac{1}{(2\pi)^k} + \frac{N}{2} \ln \frac{1}{|\Sigma|} - \frac{1}{2} \sum_{i=1}^N (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$

①

$$\Rightarrow \frac{\partial \ell}{\partial \mu_s} = \sum_{i=1}^N \left(-\frac{1}{2} \frac{\partial}{\partial \mu_s} \left(\sum_{1 \leq m, n \leq k} \Sigma_{mn}^{-1} (x_i - \mu)_m (x_i - \mu)_n \right) \right)$$

$$= \sum_{i=1}^N \left(\sum_{mn}^{-1} (x_i - \mu)_s + \sum_{m=s, n \neq s} \sum_{mn}^{-1} (x_i - \mu)_n + \sum_{m \neq s, n=s} \sum_{mn}^{-1} (x_i - \mu)_m \right)$$

$\therefore \Sigma$ is symmetric (by positive)

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

$$\Rightarrow \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \quad \#$$

$$\textcircled{2} \quad \ell(\theta) = \frac{N}{2} \ln \frac{1}{(2\pi)^k} + \frac{N}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^N \text{trace}(\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T)$$

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

$$\Rightarrow \underline{\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})(x_i - \hat{\mu})^T} \quad \#$$