BÀI TẬP CÁ NHÂN MÔN THỐNG KÊ NHIỀU CHIỀU

Tuần 2

Đinh Anh Huy - 18110103

Kết quả 4.9. Cho $\mathbf{A} \in \mathbb{R}^{k\times k}$ là ma trận đối xứng và vectơ $\mathbf{x} \in \mathbb{R}^{k\times 1}$. Khi đó:

(a) $\mathbf{x}^T\mathbf{A}\mathbf{x} = tr(\mathbf{x}^T\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{x}\mathbf{x}^T)$

(b) $tr(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$, trong đó λ_i là các trị riêng của $\mathbf{A}$.

Chứng minh

(a) Vì $\underset{(1\times k)}{\mathbf{x}^T}\underset{(k\times k)}{\mathbf{A}}\underset{(k\times 1)}{\mathbf{x}}$ là một số thực nên $\mathbf{x}^T\mathbf{A}\mathbf{x} = tr(\mathbf{x}^T\mathbf{A}\mathbf{x})$.

Ta có kết quả sau: Với hai ma trận $\mathbf{B}$, $\mathbf{C}$ bất kỳ có số chiều tương ứng là $m \times k$ và $k \times m$, ta có

$$tr(\mathbf{BC} = tr(\mathbf{CB}).$$

Thật vậy,

$$tr(\mathbf{BC}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{k} b_{ij}c_{ji}\right) = \sum_{j=1}^{k}\left(\sum_{i=1}^{m} c_{ji}b_{ij}\right) = tr(\mathbf{CB}).$$

Khi đó, với $\mathbf{B} = \mathbf{x}^T \in \mathbb{R}^{1\times k}$ và $\mathbf{C} = \mathbf{A}\mathbf{x} \in \mathbb{R}^{k\times 1}$, ta có

$$tr(\mathbf{x}^T\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{x}\mathbf{x}^T).$$

Như vậy $\mathbf{x}^T\mathbf{A}\mathbf{x} = tr(\mathbf{x}^T\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{x}\mathbf{x}^T)$.

(b) Ta có *định lý phân tích phổ (spectral decomposition)* với $\mathbf{A} \in \mathbb{R}^{k\times k}$ là ma trận đối xứng, thì

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i e_i e_i^T = \mathbf{P}^T\mathbf{\Lambda}\mathbf{P}$$

trong đó λ_i là các trị riêng của $\mathbf{A}$, e_i là các vectơ riêng chuẩn hoá của $\mathbf{A}$, $\underset{(k\times k)}{\mathbf{P}} = (e_1, ..., e_k)^T$, $\underset{(k\times k)}{\mathbf{\Lambda}} = diag(\lambda_1, \lambda_2, ..., \lambda_k)$.

Áp dụng định lý trên ta có

$$\mathbf{A} = \mathbf{P}^T\mathbf{\Lambda}\mathbf{P}$$

trong đó $\mathbf{P}\mathbf{P}^T = \mathbf{I}_k$ và $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_k)$

Khi đó

$$tr(\mathbf{A}) = tr(\mathbf{P}^T\mathbf{\Lambda}\mathbf{P}) = tr(\mathbf{\Lambda}\mathbf{P}\mathbf{P}^T) = tr(\mathbf{\Lambda}) = \sum_{i=1}^{k} \lambda_i.$$

Kết quả 4.10. Cho ma trận xác định dương $\mathbf{B} \in \mathbb{R}^{p\times p}$ và số thực dương b, khi đó

$$\frac{1}{\mathbf{\Sigma}^b} e^{-tr(\mathbf{\Sigma}^{-1}\mathbf{B})/2} \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-bp}$$

với mọi $\underset{(p\times p)}{\mathbf{\Sigma}}$ xác định dương. Dấu "=" xảy ra khi $\mathbf{\Sigma} = (1/2b)\mathbf{B}$.

Chứng minh

Với mọi $\mathbf{y} \in \mathbb{R}^p \neq 0$, ta có

$$\mathbf{y}^T \mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}\mathbf{y} = \left(\mathbf{B}^{1/2}\mathbf{y}\right)^T \mathbf{\Sigma}^{-1} \left(\mathbf{B}^{1/2}\mathbf{y}\right) > 0 \quad (\text{Do } \mathbf{\Sigma} \succ 0,\ \mathbf{B} \succ 0)$$

Do đó $\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}$ là ma trận xác định dương.

Gọi $\lambda_1, \lambda_2, ..., \lambda_p$ là các trị riêng của $\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}$. Do $\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2} \succ 0$ nên $\lambda_i > 0, i = 1, ..., p$. Vì $\mathbf{B}$ là ma trận xác định dương nên tồn tại $\mathbf{B}^{1/2}$ là ma trận căn bậc hai của $\mathbf{B}$. Khi đó, ta có

$$tr(\mathbf{\Sigma}^{-1}\mathbf{B}) = tr\left((\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2})\mathbf{B}^{1/2}\right) = tr(\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}) = \sum_{i=1}^{p} \lambda_i$$

và

$$\left|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}\right| = \prod_{i=1}^{p} \lambda_i.$$

Hơn nữa

$$\left|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}\right| = \left|\mathbf{B}^{1/2}\right| \left|\mathbf{\Sigma}^{-1}\right| \left|\mathbf{B}^{1/2}\right| = |\mathbf{\Sigma}|^{-1} |\mathbf{B}| = \frac{|\mathbf{B}|}{|\mathbf{\Sigma}|}$$

Suy ra

$$\frac{1}{|\mathbf{\Sigma}|} = \frac{|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}|}{|\mathbf{B}|} = \frac{\prod_{i=1}^{p} \lambda_i}{|\mathbf{B}|}.$$

Ta có

$$\frac{1}{|\mathbf{\Sigma}|^b} e^{-tr(\mathbf{\Sigma}^{-1}\mathbf{B})/2} = \frac{1}{|\mathbf{B}|^b} \left(\prod_{i=1}^{p} \lambda_i\right)^b e^{-\sum_{i=1}^{p} \lambda_i/2} = \frac{1}{|\mathbf{B}|^b} \prod_{i=1}^{p} \lambda_i^b e^{-\lambda_i/2}.$$

Xét hàm số $\ell(\lambda) = \lambda^b e^{-\lambda/2}$ với $\lambda, b \in \mathbb{R}, \lambda, b > 0$, ta có

$$\frac{\partial \ell}{\partial \lambda} = b\lambda^{b-1} e^{-\lambda/2} - \frac{1}{2}\lambda^b e^{-\lambda/2} = \lambda^n e^{-\lambda/2} \left(b\lambda^{-1} - \frac{1}{2}\right)$$

Giải phương trình $\dfrac{\partial \ell}{\partial \lambda} = 0$ ta được

$$\lambda^n e^{-\lambda/2} \left(b\lambda^{-1} - \frac{1}{2}\right) = 0 \Leftrightarrow b\lambda^{-1} - \frac{1}{2} = 0 \Leftrightarrow \lambda = 2b.$$

Hơn nữa

$$\begin{cases} \frac{\partial}{\partial\lambda}\ell(\lambda) > 0 & \text{với } \lambda < 2b \\ \frac{\partial}{\partial\lambda}\ell(\lambda) < 0 & \text{với } \lambda > 2b \end{cases}$$

Như vậy ℓ đạt giá trị lớn nhất tại $\lambda = 2b$ hay $\lambda^b e^{-\lambda/2} \leq (2b)^b e^{-b}$. Do đó ta có

$$\begin{aligned} \lambda_i^b e^{-\lambda_i/2} &\leq (2b)^b e^{-b} \quad \forall i = 1, ..., p \\ \prod_{i=1}^p \lambda_i^b e^{-\lambda_i/2} &\leq \prod_{i=1}^p (2b)^b e^{-b} \\ \frac{1}{|\mathbf{B}|^b} \prod_{i=1}^p \lambda_i^b e^{-\lambda_i/2} &\leq \frac{1}{|\mathbf{B}|^b} \prod_{i=1}^p (2b)^b e^{-b} = \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-pb} \\ \frac{1}{\mathbf{\Sigma}^b} e^{-tr(\mathbf{\Sigma}^{-1}\mathbf{B})/2} &\leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-bp} \end{aligned}$$

Khi đó, dấu "=" xảy ra khi $\mathbf{\Sigma} = \dfrac{1}{2b}\mathbf{B}$. Thật vậy,
Chiều thuận:
Với $\lambda = 2b$ thì

$$\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2} = \begin{bmatrix} 2b & 0 & \cdots & 0 \\ 0 & 2b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2b \end{bmatrix} = 2b\mathbf{I}_p$$

Suy ra $\mathbf{\Sigma} = \dfrac{1}{2b}\mathbf{B}$.
Chiều đảo:
Với $\mathbf{\Sigma} = \dfrac{1}{2b}\mathbf{B}$, ta có

$$\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2} = \mathbf{B}^{1/2}(2b)\mathbf{B}^{-1}\mathbf{B}^{1/2} = (2b)\mathbf{I}_p$$

và

$$tr(\mathbf{\Sigma}^{-1}\mathbf{B}) = tr(\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}) = tr((2b)\mathbf{I}_p) = 2bp$$

Hơn nữa

$$\frac{1}{|\mathbf{\Sigma}|} = \frac{|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}|}{|\mathbf{B}|} = \frac{|(2b)\mathbf{I}_p|}{|\mathbf{B}|} = \frac{(2b)^p}{|\mathbf{B}|}$$

Khi đó ta có kết quả

$$\frac{1}{\mathbf{\Sigma}^b} e^{-tr(\mathbf{\Sigma}^{-1}\mathbf{B})/2} = \frac{1}{|\mathbf{B}|^b}(2b)^{pb}e^{-bp}$$

Kết quả 4.11. Cho $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ là mẫu ngẫu nhiên kích thước n được lấy từ tổng thể có phân phối chuẩn với trung bình $\boldsymbol{\mu}$ và hiệp phương sai $\mathbf{\Sigma}$. Khi đó

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}} \quad \text{và} \quad \hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T = \frac{n-1}{n}\mathbf{S}$$

tương ứng là *ước lượng hợp lý cực đại* của $\boldsymbol{\mu}$ và $\boldsymbol{\Sigma}$. Các giá trị quan sát của chúng là $\overline{\mathbf{x}}$ và $(1/n)\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$ được gọi là *ước lượng điểm hợp lý cực đại* của $\boldsymbol{\mu}$ và $\boldsymbol{\Sigma}$.

Chứng minh

$\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ là mẫu ngẫu nhiên kích thước n được lấy từ tổng thể có phân phối chuẩn $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Khi đó $\mathbf{X}_i, i = 1, ..., n$ độc lập cùng phân phối và có hàm mật độ

$$f_{\mathbf{X}_i}(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right).$$

Ta có hàm hợp lý

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) &= \prod_{i=1}^{n} f_{X_i}(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right) \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right) \end{aligned}$$

Lấy logarit hàm hợp lý ta được

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$$

Ta có

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = -\sum_{i=1}^{n} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{x}_i)$$

Giải phương trình $\frac{\partial}{\partial \boldsymbol{\mu}} \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = 0$ ta được MLE của $\boldsymbol{\mu}$ là

$$\begin{aligned} -\sum_{i=1}^{n} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{x}_i) &= 0 \\ \sum_{i=1}^{n}(\boldsymbol{\mu} - \mathbf{x}_i) &= 0 \qquad (\text{Do } \boldsymbol{\Sigma} \succ 0) \\ n\boldsymbol{\mu} - \sum_{i=1}^{n}\mathbf{x}_i &= 0 \\ \hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i &= \overline{\mathbf{x}} \end{aligned}$$

Mặt khác, vì $(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$ là một đại lượng vô hướng nên $(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) = tr\left((\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right) = tr\left((\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\right)$. Khi đó

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n} tr\left((\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\right)$$

Ta có

$$\frac{\partial}{\partial(\boldsymbol{\Sigma}^{-1})}L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = \frac{n}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

Giải phương trình $\frac{\partial}{\partial(\boldsymbol{\Sigma}^{-1})}L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}_i) = 0$ ta được MLE của $\boldsymbol{\Sigma}$ là

$$\frac{n}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T = 0$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

Như vậy, ước lượng hợp lý cực đại (MLE) của μ và Σ là

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}} \quad \text{và} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T = \frac{n-1}{n}\mathbf{S}$$

.