High Dimensional Statistics | Prof. Dr. Podolskij Mark | Homework 4

Anton Zaitsev | 0230981826 | anton.zaitsev.001@student.uni.lu | University of Luxembourg

May 9, 2024

Exercise 3

Let $X_1, \dots, X_n \in \mathbb{R}^k$ a sequence of i.i.d. random variables following $\mathcal{N}_k(\mu, \Sigma)$ with $\mu \in \mathbb{R}^k$ known. We need to show that MLE for $\Sigma \in \mathbb{R}^{k \times k}$ is given by $\hat{\Sigma}_{\text{ML}} = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T$.

The likelihood function is given by:

$$\begin{aligned} f^n_{\mu,\Sigma}(X_1, \dots, X_n) &= \prod_{i=1}^n f_{\mu,\Sigma}(X_i) \\ &= (\det \Sigma)^{-\frac{n}{2}} (2\pi)^{-\frac{nk}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^n (X_i - \mu)^T \Sigma^{-1}(X_i - \mu)\right] \end{aligned}$$

The loglikelihood function is given by:

$$\begin{aligned} \log f^n_{\mu,\Sigma}(X_1, \dots, X_n) &= \log\left\{\prod_{i=1}^n f_{\mu,\Sigma}(X_i)\right\} \\ &= -\frac{n}{2}\log(\det \Sigma) - \frac{nk}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (X_i - \mu)^T \Sigma^{-1}(X_i - \mu) \end{aligned}$$

We look for Σ that maximizes the likelihood (loglikelihood) function:

$$\hat{\Sigma}_{\text{ML}} = \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \log f^n_{\mu,\Sigma}(X_1, \dots, X_n)$$

Let $A = \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T \in \mathbb{R}^{k\times k}$ - a positive definite matrix. We will also use the following properties of the trace operator:

1. If $A \in R : A = tr(A)$
2. $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$

Then:

$$\begin{aligned} \hat{\Sigma}_{\text{ML}} &= \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \log f^n_{\mu,\Sigma}(X_1, \dots, X_n) \\ &= \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \left\{-\frac{n}{2}\log(\det \Sigma) - \frac{nk}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (X_i - \mu)^T \Sigma^{-1}(X_i - \mu)\right\} \\ &= \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \left\{-n\log(\det \Sigma) - \text{tr}\left(\sum_{i=1}^n (X_i - \mu)^T \Sigma^{-1}(X_i - \mu)\right)\right\} \\ &= \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \left\{-n\log(\det \Sigma) - \text{tr}\left(\Sigma^{-1}\sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T\right)\right\} \\ &= \underset{\Sigma \in \mathbb{R}^{k\times k}}{\operatorname{argmax}} \left\{-n\log(\det \Sigma) - \text{tr}\left(\Sigma^{-1}A\right)\right\} \end{aligned}$$

We try to maximize function $g(\Sigma) := -n\log(\det \Sigma) - \text{tr}\left(\Sigma^{-1}A\right)$ in Σ.

Since A is positive definite almost surely, then there exists matrix B s.t. $A = BB^T$ and we define $H = B^T\Sigma^{-1}B$. Then: $\Sigma = BH^{-1}B^T$ and $\det(\Sigma) = \det(BH^{-1}B^T) = \frac{\det(BB^T)}{\det(H)} = \frac{\det(A)}{\det(H)}$ and $\text{tr}(\Sigma^{-1}A) = \text{tr}(\Sigma^{-1}BB^T) = \text{tr}(B^T\Sigma^{-1}B) = \text{tr}(H)$. Then:

$$g(\Sigma) = -n\log\left(\frac{\det(A)}{\det(H)}\right) - \text{tr}\,(H) = -n\log\left(\det(A)\right) + n\log(\det(H)) - \text{tr}\,(H)$$

The Cholesky decomposition states that any positive definite matrix can be decomposed into the product of a lower triangular matrix and its conjugate transpose. Thus, there exists a lower triangular matrix C s.t. $H = CC^T$. Then:

$$g(\Sigma) = -n\log(\det(A)) + n\log(\det(C)^2) - \operatorname{tr}\left(CC^T\right)$$

Since C is lower triangular matrix, its determinant is the product of its diagonal elements. The trace of the product CC^T is the sum of the squares of all elements of C along its main diagonal and below. Then:

$$\begin{aligned} g(\Sigma) &= -n\log(\det(A)) + n\log(\prod_{j=1}^{k} C_{jj}^2) - \sum_{j=1}^{k} C_{jj}^2 \\ &= -n\log(\det(A)) + \sum_{j=1}^{k} n\log C_{jj}^2 - \sum_{j=1}^{k} C_{jj}^2 - \sum_{i\neq j}^{k} C_{ij}^2 \\ &= -n\log(\det(A)) + \sum_{j=1}^{k}\left(n\log C_{jj}^2 - C_{jj}^2\right) - \sum_{i\neq j}^{k} C_{ij}^2 \end{aligned}$$

By maximizing above equality, we get that $C_{ij} = 0$ for $i \neq j$ and $C_{jj}^2 = n$(since $\frac{d}{dx}(n\log x - x) = 0 \iff \frac{n}{x} - 1 = 0 \iff x = n$), making C take the form:

$$C = \begin{bmatrix} \sqrt{n} & 0 & \cdots & 0 \\ 0 & \sqrt{n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{n} \end{bmatrix}$$

Then: $H = n \cdot I_k$, with $I_k - k$-dimensional identity matrix, and $\Sigma = \frac{1}{n}BB^T = \frac{1}{n}A$. Thus, $g(\Sigma)$ is maximized with $\Sigma = \frac{1}{n}A$ and $\hat{\Sigma}_{\text{ML}} = \frac{1}{n}A = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T$