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

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

May 7, 2024

Exercise 2

Let X be a k-dimensional vector with $\mathbb{E}[X] = 0 \in \mathbb{R}^k$ and $\mathbb{E}[XX^T] = \Sigma \in \mathbb{R}^{k\times k}$. The first principal component $\beta_1 \in \mathbb{R}^k$ is the eigenvector of Σ that corresponds to the largest eigenvalue λ_1 of Σ. The goal is to find vector $\beta \in \mathbb{R}^k$ s.t. $||\beta||^2 = 1, <\beta_1, \beta>= 0$ which maximizes $var(\beta^T X)$ and to deduce that β is the eigenvector of Σ that corresponds to the second largest eigenvalue λ_2 of Σ, i.e. $\beta = \beta_2$.

Notice that

$$var(\beta^T X) = \beta^T \mathbb{E}[XX^T]\beta = \beta^T \Sigma \beta$$

Since Σ is the covariance matrix, is it symmetric and positive definite. We can perform eigendecomposition of Σ:

$$\Sigma = O\Lambda O^T$$

Here, $O \in \mathbb{R}^{k\times k}$ - matrix of orthonormal (unitary in length and orthogonal to each other) eigenvectors of $\Sigma : O = [\beta_1|\beta_2|\dots|\beta_k]$
Matrix $\Lambda \in \mathbb{R}^{k\times k}$ - diagonal matrix, where each element on the diagonal is the eigenvalue of Σ s.t. $\Lambda_{11} = \lambda_1, \Lambda_{22} = \lambda_2, \dots, \Lambda_{kk} = \lambda_k, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$. Note that each column O_i is the eigenvector of λ_i. Then, we have the following problem:

$$\underset{||\beta||^2=1,<\beta,\beta_1>=0}{\text{argmax}} \{\beta^T O\Lambda O^T \beta\}$$

Note that since we seek for a vector β that is orthogonal to β_1 and unitary in length, we then essentially seek for a vector in the set $\{\beta_2, \dots, \beta_k\}$. This is due to the fact that eigenvectors in O are orthonormal and form a basis for the vector space $\mathbb{R}^{k\times k}$. The problem is then can be rewritten in the following way:

$$\underset{||\beta||^2=1,<\beta,\beta_1>=0}{\text{argmax}} \{\beta^T O\Lambda O^T \beta\} =$$

$$= \underset{<\beta_i,\beta_1>=0, i\in\{2,\dots,k\}}{\text{argmax}} \left\{ \begin{bmatrix} \beta_{i1}\beta_{i2}\dots\beta_{ik} \end{bmatrix}_{1\times k} \begin{bmatrix} \beta_{11}\dots\beta_{d1} \\ \beta_{12}\dots\beta_{d2} \\ \vdots \\ \beta_{1d}\dots\beta_{dd} \end{bmatrix}_{k\times k} \begin{bmatrix} \lambda_1 0\dots 0 \\ 0\lambda_2\dots 0 \\ \vdots \\ 00\dots\lambda_d \end{bmatrix}_{k\times k} \begin{bmatrix} \beta_{11}\dots\beta_{1d} \\ \beta_{21}\dots\beta_{2d} \\ \vdots \\ \beta_{d1}\dots\beta_{dd} \end{bmatrix}_{k\times k} \begin{bmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{ik} \end{bmatrix}_{k\times 1} \right\}$$

When we multiply vector β^T with O, we get a row vector of size $1\times k$ with 1 at position i and zeros on other positions: $\beta^T O = \begin{bmatrix} 0_1\dots 1_i\dots 0_k \end{bmatrix}_{1\times k}$. Then we scale this vector by λ_i (remember that λ_i's are ordered in the decreasing order in Λ): $\begin{bmatrix} 0_1\dots\lambda_i\dots 0_k \end{bmatrix}_{1\times k}$. For $O^T\beta$, we get a column vector of size $k\times 1$ with 1 at position i and zeros on other positions: $O^T\beta = \begin{bmatrix} 0_1 \\ \vdots \\ 1_i \\ \vdots \\ 0_k \end{bmatrix}_{k\times 1}$. Finally, we take the product of the row vector and column vector and get λ_i as a result. The largest eigenvalue for $i \in \{2,\dots,k\}$ is λ_2. Thus, vector β, which is unitary in length and orthogonal to β_1 and that maximizes $var(\beta^T X)$ is β_2 - eigenvector of Σ that corresponds to the second largest eigenvalue of Σ.

Exercise 3

Let $X_1, \ldots, X_n \in \mathbb{R}^k$ 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 $\hat{\Sigma}_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)^T$.

The likelihood function is given by:

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

The loglikelihood function is given by:

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

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

$$
\hat{\Sigma}_{\mathrm{ML}} = \operatorname*{argmax}_{\Sigma \in \mathbb{R}^{k \times k}} \log f_{\mu,\Sigma}^{n}(X_1, \ldots, X_n)
$$

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

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

Then:

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

We try to maximize function $g(\Sigma) := -n \log(\det \Sigma) - \operatorname{tr}\left(\Sigma^{-1} A\right)$ 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 $\operatorname{tr}(\Sigma^{-1} A) = \operatorname{tr}(\Sigma^{-1} BB^T) = \operatorname{tr}(B^T \Sigma^{-1} B) = \operatorname{tr}(H)$. Then:

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

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:

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

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:

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

By maximizing above equality, we get that $C_{ij} = 0$ for $i \neq j$ and $C_{jj}^2 = n$(since $\frac{d}{dx}\left(n\log x - x\right) = 0 \iff \frac{n}{x} - 1 = 0 \iff x = n$), 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 $\hat{\Sigma}_{\text{ML}} = \frac{1}{n}A = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T$