Name: Hamed Vaheb

High Dimensional Statistics: Exercise Sheet 4

Exercise 3. Let $X_1, \cdots, X_n$ be a sequence of i.i.d. random variables with common distribution $\mathcal{N}_k(\mu, \Sigma)$. Assume that the mean vector $\mu \in \mathbb{R}^k$ is known. Show that the maximum likelihood estimator of $\Sigma \in \mathbb{R}^{k \times k}$ is given by:

$$\widehat{\Sigma}_{ML} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)^T$$

Solution. For a given density function f of random variables $X_1, \cdots, X_n$, the likelihood function w.r.t parameter under study, i.e., Σ is defined as follows:

$$l_\Sigma(X_1, \cdots, X_n) = log f^n_{\mu,\Sigma}(X_1, \cdots, X_n)$$

As $X_1, \cdots, X_n \sim \mathcal{N}_k(\mu, \Sigma)$, the density function is defined as:

$$\begin{aligned}
f^n_{\mu,\Sigma}(X_1, \cdots, X_n) &= \prod_{i=1}^{n} \big\{ (det(\Sigma))^{-1/2} (2\pi)^{-k/2} exp(-\frac{1}{2}(X_i - \mu)^\intercal \Sigma^{-1} (X_i - \mu)) \big\} \\
&= (det \Sigma)^{\frac{n}{2}} (2\pi)^{\frac{kn}{2}} exp(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \mu)^\intercal \Sigma^{-1} (X_i - \mu)
\end{aligned}$$

Thus, substituting the density function in the log-likelihood function yields:

$$log f^n_{\mu,\Sigma}(X_1, \cdots, X_n) = -\frac{kn}{2} log(2\pi) - \frac{n}{2} log(det(\Sigma)) - \frac{1}{2} \sum_{i=1}^{n} (X_i - \mu)^\intercal \Sigma^{-1} (X_i - \mu)$$

The maximum likelihood estimator (MLE) $\widehat{\Sigma}$ of the parameter Σ is the following:

$$\widehat{\Sigma}_{ML} = \arg\max_{\Sigma} l_\Sigma(X_1, \cdots, X_n) = \arg\max_{\Sigma} log f^n_{\mu,\Sigma}(X_1, \cdots, X_n)$$

$$\begin{aligned}
\widehat{\Sigma}_{ML} &= \arg\max_{\Sigma} l_\Sigma(X_1, \cdots, X_n) = \arg\max_{\Sigma} log f^n_{\mu,\Sigma}(X_1, \cdots, X_n) \\
&= \arg\max_{\Sigma} \big[-\frac{kn}{2} log(2\pi) - \frac{n}{2} log(det(\Sigma)) - \frac{1}{2} \sum_{i=1}^{n} (X_i - \mu)^\intercal \Sigma^{-1} (X_i - \mu) \big]
\end{aligned}$$

We aim at finding the MLE by finding the critical points of $log f^n_{\mu,\Sigma}$. For this purpose, we define a new variable $\Sigma' = \Sigma^{-1}$ and instead of computing the derivative of the function w.r.t Σ, we do it w.r.t Σ'. We will also reformulate $log f^n_{\mu,\Sigma}$ based on $\Sigma' = \Sigma^{-1}$. This changing of variable yields the same critical point due to the following observation:

$$\max_{\Sigma} l(\Sigma) = \max_{\Sigma'} l(\Sigma')$$

In order to reformulate, we define $A(\Sigma)$ and $B(\Sigma)$ as the following:

$$\begin{aligned}
A(\Sigma) &= -\frac{n}{2} log(det(\Sigma)) \\
&= \frac{n}{2} log(det(\Sigma))^{-1} \\
&= \frac{n}{2} log(det(\Sigma^{-1}))
\end{aligned}$$

$B(\Sigma) = -\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^\intercal \Sigma^{-1}(X_i-\mu)$

We can rewrite $A(\Sigma)$ and $B(\Sigma)$ based on $\Sigma' = \Sigma^{-1}$:

$A(\Sigma') = \frac{n}{2}log(det(\Sigma'))$

$B(\Sigma') = -\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^\intercal \Sigma'(X_i-\mu)$

We now reformulate $logf^n_{\mu,\Sigma}$ as following:

$$\begin{aligned} logf^n_{\mu,\Sigma}(X_1,\cdots,X_n) &= -\frac{kn}{2}log(2\pi) + A(\Sigma) + B(\Sigma) \\ logf^n_{\mu,\Sigma'}(X_1,\cdots,X_n) &= -\frac{kn}{2}log(2\pi) + A(\Sigma') + B(\Sigma') \end{aligned}$$

We compute the derivative of log-likelihood function w.r.t Σ' as following:

$$\begin{aligned} \frac{d}{d\Sigma'}logf^n_{\mu,\Sigma'}(X_1,\cdots,X_n) &= \frac{d}{d\Sigma}\big(A(\Sigma') + B(\Sigma')\big) \\ &= \frac{d}{d\Sigma'}A(\Sigma') + \frac{d}{d\Sigma'}B(\Sigma') \end{aligned}$$

In order to compute the first term, $\frac{d}{d\Sigma'}A(\Sigma')$, we proceed as following:

$$\begin{aligned} \frac{\partial}{\partial\Sigma'_{ij}}A(\Sigma') &= \frac{\partial}{\partial\Sigma'_{ij}}\frac{n}{2}log(det(\Sigma')) \\ &= \frac{n}{2}\frac{1}{det(\Sigma')}\frac{\partial det(\Sigma')}{\partial\Sigma'_{ij}} \\ &= \frac{n}{2det(\Sigma')}adj(\Sigma')_{ij} && \text{(Jacobi's formula)} \\ &= \frac{n}{2det(\Sigma')}det(\Sigma')(\Sigma'^{-1})_{ji} && (adj(A) = det(A)A^{-1}) \\ &= \frac{n}{2}(\Sigma'^{-1})_{ji} \\ &= \frac{n}{2}(\Sigma'^{-1})_{ij} && (\Sigma \text{ and therefore } \Sigma' \text{ is symmetric}) \end{aligned}$$

The third equality is a result of Jacobi's formula, which expresses the derivative of the determinant of a matrix A in terms of the adjugate of A and the derivative of A as the following:

$$\frac{d}{dt}det(A(t)) = tr\Big(adj(A(t))\frac{dA(t)}{dt}$$

We conclude that

$$\frac{d}{d\Sigma'}A(\Sigma') = \frac{n}{2}\Sigma'^{-1} = \frac{n}{2}\Sigma$$

To compute the second term, we have:

$$\frac{d}{d\Sigma'}B(\Sigma') = \frac{d}{d\Sigma'}\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^\intercal\Sigma'(X_i-\mu)$$

We now make this observation that each term of the sum, $(X_i-\mu)^\intercal\Sigma'(X_i-\mu)$, is a scalar, hence the trace of this quantity equates itself, i.e.,

$$(X_i-\mu)^\intercal\Sigma'(X_i-\mu)=tr((X_i-\mu)^\intercal\Sigma'(X_i-\mu))$$

Consequently,

$$\sum_{i=1}^n(X_i-\mu)^\intercal\Sigma'(X_i-\mu)=\sum_{i=1}^n tr((X_i-\mu)^\intercal\Sigma'(X_i-\mu))$$

Going back to computing derivative of $B(\Sigma')$:
Consequently,

$$\begin{aligned}
\frac{d}{d\Sigma'}B(\Sigma')&=-\frac{d}{d\Sigma'}\frac{1}{2}\sum_{i=1}^n(X_i-\mu)^\intercal\Sigma'(X_i-\mu)\\
&=-\frac{1}{2}\sum_{i=1}^n\frac{d}{d\Sigma'}(X_i-\mu)^\intercal\Sigma'(X_i-\mu)\\
&=-\frac{1}{2}\sum_{i=1}^n\frac{d}{d\Sigma'}tr\big((X_i-\mu)^\intercal\Sigma'(X_i-\mu)\big)\\
&=-\frac{1}{2}\sum_{i=1}^n\frac{d}{d\Sigma'}tr\big(\underbrace{(X_i-\mu)^\intercal}_{\text{P}}\Sigma'\underbrace{(X_i-\mu)}_{\text{Q}}\big)\\
&=-\frac{1}{2}\sum_{i=1}^n\frac{d}{d\Sigma'}tr(P\Sigma'Q)\\
&=-\frac{1}{2}\sum_{i=1}^n QP && dtr(AXB)/dX=BA\\
&=-\frac{1}{2}\sum_{i=1}^n(X_i-\mu)^\intercal(X_i-\mu)
\end{aligned}$$

We feed the obtained derivatives of the two terms into the derivative of the log-likelihood function as following:

$$\begin{aligned}
\frac{d}{d\Sigma'}logf^n_{\mu,\Sigma'}(X_1,\cdots,X_n)&=\frac{d}{d\Sigma'}A(\Sigma')+\frac{d}{d\Sigma'}B(\Sigma')\\
&=\frac{n}{2}\Sigma-\frac{1}{2}\sum_{i=1}^n(X_i-\mu)^\intercal(X_i-\mu)
\end{aligned}$$

In order to find the critical points of the log-likelihood function, we set its derivative to zero:

$$\frac{n}{2}\Sigma-\frac{1}{2}\sum_{i=1}^n(X_i-\mu)^\intercal(X_i-\mu)=0\implies\Sigma=\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^\intercal(X_i-\mu)$$

As the log-likelihood is a concave function, we conclude that this critical point is minimum and hence we have obtained the MLE.