MLE FOR MULTIVARIATE NORMAL

Given \( x_1, \ldots, x_n \sim N(\mu, \Sigma) \) where \( N(\mu, \Sigma) \) represents the \( p \)-dimensional multivariate normal with mean vector \( \mu \in \mathbb{R}^p \) and covariance matrix \( \Sigma \in \mathbb{R}^{p \times p} \), we want to compute the maximum likelihood estimator \( \hat{\mu} \) and \( \hat{\Sigma} \). The computation is as simple as the MLE for univariate normal if you know how to compute the derivative with respect to a matrix, etc.

The log-likelihood is

\[
L(\mu, \Sigma) = \sum_{i=1}^{n} \log \left( |2\pi \Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right] \right) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu).
\]

We have

\[
\frac{\partial L(\mu, \Sigma)}{\partial \mu} = \frac{1}{2} \sum_{i=1}^{n} 2\Sigma^{-1} (x_i - \mu) = 0
\]

which implies

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

Furthermore, we have

\[
\frac{\partial L(\mu, \Sigma)}{\partial \Sigma} = -\frac{n}{2} \frac{\partial \log |\Sigma|}{\Sigma} - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}{\partial \Sigma}
\]

\[
= -\frac{n}{2} \Sigma^{-1} - \frac{1}{2} \sum_{i=1}^{n} -\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T \Sigma^{-1}
\]

\[
= 0
\]

where we used the matrix differentiation rules

\[
\frac{d \log |\Sigma|}{\Sigma} = \frac{1}{|\Sigma|} \frac{d |\Sigma|}{d \Sigma} = \frac{1}{|\Sigma|} |\Sigma|^{-1} = \Sigma^{-1}
\]

and

\[
\frac{d a^T \Sigma^{-1} b}{d \Sigma} = -\Sigma^{-1} b a^T \Sigma^{-1}.
\]

It is easy to see that \( \frac{\partial L(\mu, \Sigma)}{\partial \Sigma} = 0 \) leads to

\[
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T.
\]

This establishes that the sample mean and sample covariance are the MLEs.