Maximum Likelihood Estimation for the Mean $\boldsymbol{\mu}$ in a Multivariate Gaussian Model

To estimate the mean $\boldsymbol{\mu}$ of a multivariate Gaussian model using Maximum Likelihood Estimation (MLE), given a set of sampled data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, the following steps can be followed:

Write the Likelihood Function

The probability density function (PDF) for a multivariate Gaussian distribution is:

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

Where:

- $\mathbf{x} \in \mathbb{R}^d$ is a d-dimensional random vector,
- $\boldsymbol{\mu} \in \mathbb{R}^d$ is the mean vector,
- $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is the covariance matrix.

Given n i.i.d. samples $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, the likelihood function is the product of the individual PDFs of all the data points:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Taking the logarithm of the likelihood function gives the log-likelihood:

$$\log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \log p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Substituting the PDF into the log-likelihood:

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

Maximize the Log-Likelihood with Respect to $\boldsymbol{\mu}$

To find the maximum likelihood estimate (MLE) for $\boldsymbol{\mu}$, we take the derivative of the log-likelihood function with respect to $\boldsymbol{\mu}$ and set it equal to zero. The term that involves $\boldsymbol{\mu}$ is:

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

Taking the derivative of this term with respect to $\boldsymbol{\mu}$:

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

Set this derivative equal to zero to maximize the log-likelihood:

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

This simplifies to:

$$\boldsymbol{\Sigma}^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i\right) = n\boldsymbol{\mu}$$

Solving for $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

Final Answer

Thus, the Maximum Likelihood Estimate (MLE) for the mean vector $\boldsymbol{\mu}$ is simply the sample mean of the data:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

This is the arithmetic average of the data points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.