Problem

Given the multivariate normal probability density function:

$$f(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right),$$

where $x, \mu \in \mathbb{R}^k$, Σ is a $k \times k$ positive definite matrix, and $|\Sigma|$ denotes the determinant of Σ.

Show that:

$$\int_{\mathbb{R}^k} f(x)\, dx = 1$$

Solution

Step 1: Change of Variables

We apply the change of variables:

$$z = \Sigma^{-1/2}(x-\mu) \quad \Rightarrow \quad x = \mu + \Sigma^{1/2} z$$

The Jacobian determinant of this linear transformation is:

$$\left| \frac{dx}{dz} \right| = |\Sigma^{1/2}| = \sqrt{|\Sigma|}$$

Step 2: Rewrite the Integral

Under this change of variables:

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = z^T z = \|z\|^2$$
$$dx = \sqrt{|\Sigma|}\, dz$$

The integral becomes:

$$\int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2} z^T z\right) \cdot \sqrt{|\Sigma|}\, dz$$

Simplifying:

$$= \int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|z\|^2\right) dz$$

Step 3: Recognize the Standard Normal Density

The integrand is the probability density function of a standard k-dimensional multivariate normal distribution:

$$Z \sim \mathcal{N}(0, I)$$

It is well-known that:

$$\int_{\mathbb{R}^k} \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|z\|^2\right) dz = 1$$

Conclusion

Therefore, the original multivariate normal density is properly normalized:

$$\int_{\mathbb{R}^k} f(x)\, dx = 1$$

Matrix Calculus and Multivariate Gaussian MLE

(a) Show that $\frac{\partial}{\partial A}\operatorname{tr}(AB) = B^T$

We start by expanding the trace using matrix elements:

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}b_{ki}$$

Now take the partial derivative with respect to a_{pq}:

$$\frac{\partial}{\partial a_{pq}}\operatorname{tr}(AB) = \sum_{i=1}^{n}\sum_{k=1}^{n} b_{ki}\cdot\frac{\partial a_{ik}}{\partial a_{pq}}$$

Using the Kronecker delta:

$$\frac{\partial a_{ik}}{\partial a_{pq}} = \delta_{ip}\delta_{kq} \Rightarrow \text{Only when } i = p, k = q \text{ is the term 1, else 0}$$

Thus only one term survives:

$$\frac{\partial}{\partial a_{pq}}\operatorname{tr}(AB) = b_{qp}$$

This is the (p,q)-th entry of B^T, hence:

$$\frac{\partial}{\partial A}\operatorname{tr}(AB) = B^T$$

(b) Show that $x^TAx = \operatorname{tr}(xx^TA)$

Let $x \in \mathbb{R}^n$ be a column vector and $A \in \mathbb{R}^{n\times n}$ a square matrix.

Using the cyclic property of the trace:

$$\operatorname{tr}(xx^TA) = \operatorname{tr}(Axx^T) = \operatorname{tr}((Ax)x^T)$$

Now recall:

$$\operatorname{tr}((Ax)x^T) = \sum_{i=1}^{n}(Ax)_i x_i = x^TAx$$

Therefore:

$$x^TAx = \operatorname{tr}(xx^TA)$$

(c) Derive the Maximum Likelihood Estimators for the Multivariate Gaussian

Assume we have n samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \in \mathbb{R}^d$, each drawn i.i.d. from a multivariate normal distribution:

$$\boldsymbol{x}^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

The probability density function is:

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

The log-likelihood over n samples is:

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

To maximize with respect to $\boldsymbol{\mu}$, take derivative and set to 0:

$$\frac{\partial \log L}{\partial \boldsymbol{\mu}} = \Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}^{(i)} - \boldsymbol{\mu}) = 0 \Rightarrow \hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}^{(i)}$$

Plug $\hat{\boldsymbol{\mu}}$ back into the log-likelihood and take derivative w.r.t. Σ, we get:

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

Therefore, the MLEs are:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}^{(i)}, \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}^{(i)} - \hat{\boldsymbol{\mu}})(\boldsymbol{x}^{(i)} - \hat{\boldsymbol{\mu}})^T$$