Assignment 5

November 25, 2025

1 Normalization of Multivariate Gaussian Distribution

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

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

where $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^k$ and Σ is a $k \times k$ positive definite matrix. We need to show that $\int_{\mathbb{R}^k} f(\mathbf{x})\, d\mathbf{x} = 1$.

Diagonalization and Change of Variables

Since Σ is a positive definite matrix, it can be orthogonally diagonalized:

$$\Sigma = \mathbf{Q}\mathbf{D}\mathbf{Q}^T,$$

where $\mathbf{Q}$ is an orthogonal matrix ($\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$) and $\mathbf{D}$ is a diagonal matrix with positive eigenvalues. We define the square root of Σ as $\Sigma^{1/2} = \mathbf{Q}\mathbf{D}^{1/2}\mathbf{Q}^T$, such that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$. The determinant relationship is $\det(\Sigma) = \det(\Sigma^{1/2})^2$, so $\det(\Sigma^{1/2}) = \sqrt{|\Sigma|}$.

We perform a change of variable:

$$\mathbf{y} = \Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) \quad \Longrightarrow \quad \mathbf{x} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{y}.$$

The **Jacobian determinant** of this transformation is $\left|\det(\frac{\partial \mathbf{x}}{\partial \mathbf{y}})\right| = \left|\det(\Sigma^{1/2})\right| = |\Sigma|^{1/2}$. Thus, $d\mathbf{x} = |\Sigma|^{1/2}\, d\mathbf{y}$.

Integral Evaluation

Substitute the change of variables into the integral. Note that:

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\Sigma^{1/2}\mathbf{y})^T \Sigma^{-1} (\Sigma^{1/2}\mathbf{y}) = \mathbf{y}^T (\Sigma^{1/2})^T \Sigma^{-1} \Sigma^{1/2} \mathbf{y} = \mathbf{y}^T\mathbf{y}.$$

The integral becomes:

$$\begin{aligned}
\int_{\mathbb{R}^k} f(\mathbf{x})\, d\mathbf{x} &= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2} \mathbf{y}^T\mathbf{y} \right) |\Sigma|^{1/2}\, d\mathbf{y} \\
&= \frac{|\Sigma|^{1/2}}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} e^{-\frac{1}{2}\sum_{i=1}^k y_i^2}\, d\mathbf{y} \\
&= \frac{1}{(\sqrt{2\pi})^k} \prod_{i=1}^k \int_{-\infty}^{\infty} e^{-y_i^2/2}\, dy_i \quad \text{(Separable integral)} \\
&= \frac{1}{(\sqrt{2\pi})^k} \left(\int_{-\infty}^{\infty} e^{-y^2/2}\, dy \right)^k
\end{aligned}$$

Using the standard Gaussian integral result from calculus, $\int_{-\infty}^{\infty} e^{-y^2/2}\, dy = \sqrt{2\pi}$, the expression simplifies to:

$$\int_{\mathbb{R}^k} f(\mathbf{x})\, d\mathbf{x} = \frac{1}{(\sqrt{2\pi})^k} \left(\sqrt{2\pi}\right)^k = 1.$$

Thus, $\boxed{\int_{\mathbb{R}^k} f(\mathbf{x})\, d\mathbf{x} = 1.}$

2 Matrix Calculus and Trace Identities

2.(1) Show that $\frac{\partial}{\partial \mathbf{A}} \operatorname{tr}(\mathbf{AB}) = \mathbf{B}^T$

The trace is defined as $\operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^{n} (\mathbf{AB})_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki}$. We calculate the derivative with respect to a specific element A_{lm}:

$$\frac{\partial}{\partial \mathbf{A}_{lm}} \operatorname{tr}(\mathbf{AB}) = \frac{\partial}{\partial \mathbf{A}_{lm}} \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki}.$$

The derivative $\frac{\partial A_{ik}}{\partial A_{lm}}$ is 1 only when $i = l$ and $k = m$, otherwise it's 0.

$$\frac{\partial}{\partial \mathbf{A}_{lm}} \operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^{n} \sum_{k=1}^{n} (\frac{\partial A_{ik}}{\partial A_{lm}}) B_{ki} = B_{ml}.$$

In matrix form, $\left[\frac{\partial}{\partial \mathbf{A}} \operatorname{tr}(\mathbf{AB})\right]_{lm} = B_{ml}$, which is the (l, m) element of $\mathbf{B}^T$. Therefore, $\boxed{\frac{\partial}{\partial \mathbf{A}} \operatorname{tr}(\mathbf{AB}) = \mathbf{B}^T.}$

2.(2) Show that $\mathbf{x}^T \mathbf{A} \mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{x}^T \mathbf{A})$

Let $\mathbf{x}$ be $n \times 1$ and $\mathbf{A}$ be $n \times n$.

Left Hand Side (LHS)

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^{n} x_i (\mathbf{A}\mathbf{x})_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j.$$

Right Hand Side (RHS)

Let $\mathbf{C} = \mathbf{x}\mathbf{x}^T \mathbf{A}$. $\mathbf{C}$ is $n \times n$. The (k, j) element of $\mathbf{x}\mathbf{x}^T$ is $(\mathbf{x}\mathbf{x}^T)_{kj} = x_k x_j$.

$$\operatorname{tr}(\mathbf{x}\mathbf{x}^T \mathbf{A}) = \sum_{k=1}^{n} (\mathbf{x}\mathbf{x}^T \mathbf{A})_{kk} = \sum_{k=1}^{n} \sum_{j=1}^{n} (\mathbf{x}\mathbf{x}^T)_{kj} A_{jk} = \sum_{k=1}^{n} \sum_{j=1}^{n} (x_k x_j) A_{jk} = \sum_{k=1}^{n} \sum_{j=1}^{n} x_k A_{jk} x_j.$$

By swapping the index variables $i \leftrightarrow k$ in the LHS and noting that the summation is identical, we conclude $\mathbf{x}^T \mathbf{A} \mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{x}^T \mathbf{A})$.

2.(3) Derive the Maximum Likelihood Estimators (MLE) for a Multivariate Gaussian

The log-likelihood function for N i.i.d. samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from $f(\mathbf{x})$ is:

$$\ell(\boldsymbol{\mu}, \Sigma) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

MLE of $\boldsymbol{\mu}$

We ignore constant terms and take the derivative with respect to $\boldsymbol{\mu}$:

$$\ell(\boldsymbol{\mu}) = -\frac{1}{2}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) + C_1.$$

Using the identity $\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x}$ and $\frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T\mathbf{x} = \mathbf{b}$, and letting $\mathbf{y}_i = \mathbf{x}_i - \boldsymbol{\mu}$:

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\mu}} &= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}}\left[\mathbf{x}_i^T\Sigma^{-1}\mathbf{x}_i - 2\mathbf{x}_i^T\Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}\right] \\ &= -\frac{1}{2}\sum_{i=1}^{N}\left[0 - 2(\Sigma^{-1}\mathbf{x}_i) + 2\Sigma^{-1}\boldsymbol{\mu}\right] \\ &= \sum_{i=1}^{N}(\Sigma^{-1}\mathbf{x}_i - \Sigma^{-1}\boldsymbol{\mu}) = \Sigma^{-1}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu}).\end{aligned}$$

Setting $\frac{\partial \ell}{\partial \boldsymbol{\mu}} = 0$:

$$\Sigma^{-1}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu}) = 0 \quad \Longrightarrow \quad \sum_{i=1}^{N}\mathbf{x}_i = \sum_{i=1}^{N}\boldsymbol{\mu} = N\boldsymbol{\mu}.$$

Thus, the MLE for the mean is the sample mean:

$$\boxed{\hat{\boldsymbol{\mu}}_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i}.$$

MLE of Σ

We substitute $\hat{\boldsymbol{\mu}}$ into the log-likelihood and maximize with respect to Σ.

$$\begin{aligned}\frac{\partial \ell}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma}\left[-\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T\Sigma^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})\right] \\ &= -\frac{N}{2}\frac{\partial \ln|\Sigma|}{\partial \Sigma} - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \Sigma}\left(\operatorname{tr}\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T\Sigma^{-1}\right)\right)\end{aligned}$$

We use the matrix calculus identities (assuming Σ is symmetric):

- $\frac{\partial \ln|\Sigma|}{\partial \Sigma} = \Sigma^{-1}$
- $\frac{\partial \operatorname{tr}(\mathbf{A}\Sigma^{-1})}{\partial \Sigma} = -\Sigma^{-1}\mathbf{A}^T\Sigma^{-1} = -\Sigma^{-1}\mathbf{A}\Sigma^{-1}$ (since $\mathbf{A}$ is symmetric here)

Let $\mathbf{A}_i = (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$.

$$\frac{\partial \ell}{\partial \Sigma} = -\frac{N}{2}\Sigma^{-1} - \frac{1}{2}\sum_{i=1}^{N}\left(-\Sigma^{-1}\mathbf{A}_i\Sigma^{-1}\right) = -\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{N}\Sigma^{-1}\mathbf{A}_i\Sigma^{-1}.$$

Setting $\frac{\partial \ell}{\partial \Sigma} = 0$:

$$\Sigma^{-1}\left(-\frac{N}{2}\mathbf{I} + \frac{1}{2}\sum_{i=1}^{N}\mathbf{A}_i\right)\Sigma^{-1} = 0.$$

Since Σ^{-1} is invertible, we must have:

$$\sum_{i=1}^{N}\mathbf{A}_i = N\mathbf{I} \quad \Longrightarrow \quad \sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T = N\Sigma.$$

Thus, the MLE for the covariance matrix is the sample covariance matrix (using N in the denominator):

$$\boxed{\widehat{\Sigma}_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T}.$$

3 Unanswered Question

What may happen if the real-world data we used doesn't satisfy normal distribution? What is the result?