Week 5 Written Assignment

Name: HanYunChen Student ID: 112652010

October 7, 2025

1. Normalization of the Multivariate Gaussian

Let

$$f(x) = \frac{1}{\sqrt{(2\pi)^k\,|\Sigma|}} \exp\Big(-\tfrac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\Big), \tag{1}$$

where $x, \mu \in \mathbb{R}^k$ and $\Sigma \in \mathbb{R}^{k\times k}$ is symmetric positive definite. We show $\int_{\mathbb{R}^k} f(x)\,dx = 1$.

Proof. Because Σ is positive definite, it has a symmetric square root $\Sigma^{1/2}$ with $(\Sigma^{1/2})^2 = \Sigma$ and $|\Sigma^{1/2}| = |\Sigma|^{1/2}$. Set the change of variables

$$y = \Sigma^{-1/2}(x-\mu) \qquad \Longleftrightarrow \qquad x = \Sigma^{1/2}y + \mu. \tag{2}$$

Then the Jacobian determinant is $dx = |\Sigma^{1/2}|\,dy = |\Sigma|^{1/2}dy$, and

$$\int_{\mathbb{R}^k} f(x)\,dx = \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\tfrac{1}{2}y^\top y\Big)\,|\Sigma|^{1/2}dy \tag{3}$$

$$= \prod_{i=1}^{k}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y_i^2/2}\,dy_i = 1, \tag{4}$$

since each 1D integral is the standard normal integral. □

2. Matrix calculus identities

Let $A, B \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$.

(a) $\frac{\partial}{\partial A}\,\mathrm{tr}(AB) = B^\top$

Using $\mathrm{tr}(AB) = \sum_{i,j} A_{ij}B_{ji}$, for a fixed entry A_{pq} we have

$$\frac{\partial}{\partial A_{pq}}\,\mathrm{tr}(AB) = B_{qp}. \tag{5}$$

Therefore the matrix of partial derivatives equals $B^\top$. □

(b) $x^\top Ax = \mathrm{tr}\big(xx^\top A\big)$

By cyclicity of trace,

$$\mathrm{tr}\Big(xx^\top A\Big) = \mathrm{tr}\Big(Axx^\top\Big) = \sum_{i,j} A_{ij}x_jx_i = x^\top Ax. \tag{6}$$

□

(c) Maximum likelihood estimators for $\mathcal{N}(\mu, \Sigma)$

Suppose $x_1, \ldots, x_n \in \mathbb{R}^k$ are i.i.d. samples from $\mathcal{N}(\mu, \Sigma)$ with density in Part 1. The log-likelihood is

$$\ell(\mu, \Sigma) = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log|\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu). \tag{7}$$

Estimator of μ. Using the identity $\nabla_\mu\big((x-\mu)^\top \Sigma^{-1}(x-\mu)\big) = -2\Sigma^{-1}(x-\mu)$,

$$\nabla_\mu \ell(\mu, \Sigma) = \sum_{i=1}^{n} \Sigma^{-1}(x_i - \mu). \tag{8}$$

Setting it to zero gives

$$\boxed{\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}. \tag{9}$$

Estimator of Σ. It is convenient to differentiate w.r.t. Σ^{-1} (denote $S = \Sigma^{-1}$). Using $\partial \log|\Sigma| / \partial \Sigma^{-1} = -\Sigma$ and

$$\frac{\partial}{\partial S} \sum_{i=1}^{n} (x_i - \mu)^\top S (x_i - \mu) = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^\top, \tag{10}$$

we get

$$\frac{\partial \ell}{\partial S} = \frac{n}{2}\Sigma - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^\top. \tag{11}$$

Setting to zero and plugging $\hat{\mu}$ yields

$$\boxed{\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^\top}. \tag{12}$$

(Note: this is the MLE. The unbiased sample covariance uses $1/(n-1)$.)

3. Unanswered questions (for reflection)

- Why are multivariate Gaussian contours ellipsoids determined by eigenvectors/eigenvalues of Σ?
- What breaks if Σ is not positive definite?
- Why does MLE use $1/n$ while the unbiased estimator uses $1/(n-1)$?

References.

- Course lecture notes and assignment instructions.
- OpenAI. (2025)ChatGPT (GPT-5) From https://chat.openai.com/