ML Week4 Assignment

WANG TZU YI

October 5, 2025

1. Proof that Multivariate Gaussian Integrates to 1

Let $x \in \mathbb{R}^d$ be a random vector following a multivariate normal distribution:

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

with mean vector $\mu \in \mathbb{R}^d$ and positive-definite covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. The probability density function (PDF) is given by:

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

Goal

Show that $f(x)$ is a valid probability density function, i.e.:

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

Change of Variables

Define the change of variables:

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

Here, $\Sigma^{1/2}$ denotes the matrix square root of Σ, satisfying $\Sigma^{1/2}(\Sigma^{1/2})^\top = \Sigma$. Because Σ is symmetric and positive definite, such a square root exists and is invertible.

Jacobian Determinant

The Jacobian matrix of the transformation is:

$$\frac{\partial x}{\partial z} = \Sigma^{1/2} \quad \Rightarrow \quad \left| \frac{\partial x}{\partial z} \right| = |\Sigma^{1/2}| = |\Sigma|^{1/2}$$

Substituting into the Integral

We now change variables:

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}z^\top z\right) |\Sigma|^{1/2} dz$$

The $|\Sigma|^{1/2}$ terms cancel:

$$= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}z^\top z\right) dz$$

This is the standard multivariate normal density $\mathcal{N}(0, I)$, which is known to integrate to 1:

$$\int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}z^\top z\right) dz = 1$$

Conclusion

Therefore, the multivariate normal distribution is a valid probability density function:

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

2. Problem (a): $\frac{\partial}{\partial A}\operatorname{tr}(AB) = B^\top$

Solution

Using standard matrix calculus identity:

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

Verification

Expanding the trace:

$$\operatorname{tr}(AB) = \sum_{i,j} A_{ij} B_{ji} \quad \Rightarrow \quad \frac{\partial \operatorname{tr}(AB)}{\partial A_{ij}} = B_{ji} \Rightarrow \frac{\partial \operatorname{tr}(AB)}{\partial A} = B^\top$$

---

3. Problem (b): $x^\top A x = \operatorname{tr}(x x^\top A)$

Proof

We use the cyclic property of the trace:

$$x^\top A x = \operatorname{tr}(x^\top A x) = \operatorname{tr}(x x^\top A)$$

Hence:

$$\boxed{x^\top A x = \operatorname{tr}(x x^\top A)}$$

---

4. Problem (c): MLE for Multivariate Gaussian Parameters

Setting

Given observations $x_1, \ldots, x_n \in \mathbb{R}^d$, assume:

$$x_i \sim \mathcal{N}(\mu, \Sigma)$$

Log-likelihood function

$$\log L(\mu, \Sigma) = -\frac{nd}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu)$$

Step 1: Derivative w.r.t. μ

$$\frac{\partial}{\partial \mu}\left[\sum_{i=1}^{n}(x_i-\mu)^\top \Sigma^{-1}(x_i-\mu)\right] = -2\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu) \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Step 2: Derivative w.r.t. Σ

Rewriting the quadratic term using trace:

$$\sum_{i=1}^{n}(x_i-\mu)^\top \Sigma^{-1}(x_i-\mu) = \mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^\top\right)$$

Thus, the MLE of Σ is:

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

Final Result (MLE Estimates)

$$\boxed{\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \hat{\Sigma} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})(x_i-\hat{\mu})^\top \end{aligned}}$$