

# MACHINE LEARNING

## Homework Week 2

September 22, 2021

Nguyen Son Tung

---

1. Proof: Multivariate Gaussian distribution normalization

### Answer

For a D-dimensional vector  $x$ , the multivariate Gaussian distribution takes the form

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

where  $\mu$  is a D-dimensional mean vector,  $\Sigma$  is a  $D \times D$  covariance matrix, and  $|\Sigma|$  denotes the determinant of  $\Sigma$

Set

$$\Delta^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

We have:

A square matrix  $A \in R^{n \times n}$  can be factored into

$$A = PDP^{-1} \quad (1)$$

Where  $A$  is symmetric its eigenvalues will be real and its eigenvectors form an orthonormal set. While  $P$  is created from the eigenvector of  $A$

$$\Rightarrow PP^T = I \text{ or } P^T = P^{-1}$$

Equation (1) can be rewritten as

$$A = PDP^T$$

Where

$$P = \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix}$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \vdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

$$\Rightarrow A = \Sigma = \sum_{i=1}^D u_i \lambda_i u_i^T \text{ or } \Sigma = \sum_{i=1}^D \lambda_i u_i u_i^T$$

Then,

$$\begin{aligned} \Sigma^{-1} &= (PDP^T)^{-1} \\ &= P^{-T} D^{-1} P^{-1} \\ &= PD^{-1}P^T \end{aligned}$$

Since D is a diagonal matrix, we have:

$$\begin{aligned} D^k &= \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \vdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n^k \end{bmatrix} \\ \Rightarrow D^{-1} &= \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \vdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{\lambda_n} \end{bmatrix} \\ \Rightarrow \Sigma^{-1} &= \sum_{i=1}^D u_i \frac{1}{\lambda_i} u_i^T \text{ or } \Sigma^{-1} = \sum_{i=1}^D \frac{1}{\lambda_i} u_i u_i^T \end{aligned}$$

So that,

$$\begin{aligned} \Delta^2 &= (x - \mu)^T \Sigma^{-1} (x - \mu) \\ &= \sum_{i=1}^D \frac{1}{\lambda_i} (x - \mu)^T u_i u_i^T (x - \mu) \\ &= \sum_{i=1}^D \frac{y_i^2}{\lambda_i}, \text{ with } y_i = u_i^T (x - \mu) \end{aligned}$$

Next, to calculate  $|\Sigma|^{1/2}$ , we have known that the determinant of a matrix is equal to the product of its eigenvalues

$$\Rightarrow |\Sigma|^{1/2} = \prod_{j=1}^D \lambda_j^{1/2}$$

So, the multivariate Gaussian distribution can be rewritten as

$$p(y) = \prod_{j=1}^D \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\}$$

$$\Rightarrow \int_{-\infty}^{\infty} p(y) dy = \prod_{j=1}^D \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left\{\frac{y_j^2}{2\lambda_j}\right\} dy_j$$

With each of j, the right equation can be represented as an univariate Gaussian distribution, since the univariate Gaussian distribution is normalized

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda)^{1/2}} \exp\left\{\frac{y^2}{2\lambda}\right\} dy = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} p(y) dy = \prod_{j=1}^D \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left\{\frac{y_j^2}{2\lambda_j}\right\} dy_j = 1$$

Therefore, the multivariate Gaussian distribution is normalized.

## 2. Conditional normal distribution

### Answer

Let:

$$\begin{aligned} \Delta^2 &= -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \\ &= -\frac{1}{2}(x^T - \mu^T) \Sigma^{-1}(x - \mu) \\ &= -\frac{1}{2}x^T \Sigma^{-1}x + \frac{1}{2}(x^T \Sigma^{-1}\mu + \mu^T \Sigma^{-1}x) - \frac{1}{2}\mu^T \Sigma^{-1}\mu \end{aligned} \quad (2)$$

Where:

- (\*)  $x$  is a  $D \times 1$  matrix  $\rightarrow x^T$  is a  $1 \times D$  matrix
- (\*)  $\mu$  is a  $D \times 1$  matrix
- (\*)  $\Sigma^{-1}$  is a  $D \times D$  covariance matrix which positive definite and symmetric

So, the dimension of

$$\begin{aligned} x^T \Sigma^{-1} \mu &= 1 \times D \otimes D \times D \otimes D \times 1 \\ &= 1 \times 1 \end{aligned}$$

$\rightarrow x^T \Sigma^{-1} \mu$  equal a numeric value

$$\begin{aligned} \Rightarrow x^T \Sigma^{-1} \mu &= (x^T \Sigma^{-1} \mu)^T \\ &= \mu^T \Sigma^{-1} x \end{aligned}$$

Therefore, equation (2) can be rewritten as

$$\Delta^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + \text{const}$$

Suppose  $x$  is a  $D$ -dimensional vector with Gaussian distribution  $\mathcal{N}(x|\mu, \Sigma)$  and that we partition  $x$  into two disjoint subsets  $x_a$  and  $x_b$

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$$

We also define corresponding partitions of the mean vector  $\mu$  given by

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$$

and of the covariance matrix  $\Sigma$  given by

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Let

$$A = \Sigma^{-1} = \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix}$$

We are looking for conditional distribution  $p(x_a|x_b)$ . We have:

$$\begin{aligned} \Delta^2 &= -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \\ &= -\frac{1}{2}(x - \mu)^T A (x - \mu) \\ &= -\frac{1}{2} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix}^T \begin{bmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{bmatrix} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} \\ &= -\frac{1}{2}(x_a - \mu_a)^T A_{aa} (x_a - \mu_a) - \frac{1}{2}(x_a - \mu_a)^T A_{ab} (x_b - \mu_b) \\ &\quad - \frac{1}{2}(x_b - \mu_b)^T A_{ba} (x_a - \mu_a) - \frac{1}{2}(x_b - \mu_b)^T A_{bb} (x_b - \mu_b) \\ &= -\frac{1}{2}x_a^T A_{aa} x_a + x_a^T [A_{aa}\mu_a - A_{ab}(x_b - \mu_b)] + \text{const} \end{aligned}$$

Compare with Gaussian distribution

$$\Delta^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + \text{const}$$

$$\begin{aligned}
&\rightarrow \begin{cases} -\frac{1}{2}x^T \Sigma^{-1}x = -\frac{1}{2}x_a^T A_{aa}x_a \\ x^T \Sigma^{-1}\mu = x_a^T [A_{aa}\mu_a - A_{ab}(x_b - \mu_b)] \end{cases} \\
&\Leftrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \Sigma^{-1}\mu = A_{aa}\mu_a - A_{ab}(x_b - \mu_b) \end{cases} \\
&\Leftrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \mu = \mu_a - A_{aa}^{-1}A_{ab}(x_b - \mu_b) \end{cases}
\end{aligned}$$

By using Schur complement,

$$\begin{aligned}
\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} &= \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1} \end{pmatrix}, \text{ with } M = (A - BD^{-1}C)^{-1} \\
&\Rightarrow \begin{cases} A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} \\ A_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \end{cases}
\end{aligned}$$

As a result,

$$\begin{aligned}
&\begin{cases} \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \\ \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{cases} \\
&\Rightarrow p(x_a|x_b) = \mathcal{N}(x_a|b, \Sigma_{a|b})
\end{aligned}$$

### 3. Marginal normal distribution

#### Answer

The marginal distribution given by

$$p(x_a) = \int p(x_a, x_b) dx_b$$

We need to integrate out  $x_b$  by looking the quadratic form related to  $x_b$

$$\begin{aligned}
\Delta^2 &= -\frac{1}{2}(x - \mu)^T A(x - \mu) \\
&= -\frac{1}{2}x_b^T A_{bb}x_b + x_b^T m + \text{const} \quad (\text{with } m = A_{bb}\mu_b - A_{ba}(x_a - \mu_a)) \\
&= -\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m) + \frac{1}{2}m^T A_{bb}^{-1}m
\end{aligned}$$

We can integrate over unnormalized Gaussian

$$\int \exp\left\{-\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m)\right\} dx_b$$

The remaining term

$$-\frac{1}{2}x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})x_a + x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})^{-1}\mu_a + \text{const}$$

Similarly we have

$$\mathbb{E}[x_a] = \mu_a$$

$$\text{cov}[x_a] = \Sigma_{aa}$$

$$\Rightarrow p(x_a) = \mathcal{N}(x_a|\mu_a, \Sigma_{aa})$$