Machine Learning 1 - Week 2

Nguyễn Thị Kiều Nhung - 11203041

August, 2022

Problem 2. Calculate

(a) The conditional of Gaussian Distribution

Quadratic form of Gaussian distribution, Let:

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

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

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

Where:

x is a $D \times 1$ matrix $\Longrightarrow$ x^T is a $1 \times D$ matrix
μ is a $D \times 1$ matrix
Σ^{-1} is a $D \times D$ covariance matrix which is positive definite and symmetric
So:

$$x^T\Sigma^{-1}\mu = (1\times D)\bigotimes(D\times D)\bigotimes(D\times 1)$$
$$= 1\times 1$$

$\Longrightarrow$ x$^T\Sigma^{-1}\mu$ likes a numeric value
$\Longrightarrow x^T\Sigma^{-1}\mu = (x^T\Sigma^{-1}\mu)^T = \mu^T\Sigma^{-1}x$

Therefore, we can rewrite:

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

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 μ given by:

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

and of the covariance matrix Σ given by:

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

Let:

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

Σ is symmetric so Σ_{aa} and Σ_{bb} are symmetric while

$$\Sigma_{ab} = \Sigma_{ba}^T$$

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

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

$$= -\frac{1}{2}(x-\mu)^T A(x-\mu) \tag{3}$$

$$= -\frac{1}{2}\begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix}^T \begin{bmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{bmatrix}\begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} \tag{4}$$

$$= -\frac{1}{2}(x_a-\mu_a)^T A_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T A_{ab}(x_b-\mu_b) - \frac{1}{2}(x_b-\mu_b)^T A_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T A_{bb}(x_b-\mu_b) \tag{5}$$

$$= -\frac{1}{2}x_a^T A_{aa} x_a + x_a^T[A_{aa}\mu_a - A_{ab}(x_b-\mu_b)] + const \tag{6}$$

Compare with Gaussian distribution:

$$\Delta^2 = -\frac{1}{2}x^T \Sigma^{-1} x + x^T \Sigma^{-1}\mu + const \tag{7}$$

$$\Longrightarrow \begin{cases} -\frac{1}{2}x^T\Sigma^{-1}x = -\frac{1}{2}x_a^T A_{aa} x_a \\ x^T \Sigma^{-1}\mu = x_a^T[A^{aa}\mu_a - A_{ab}(x_b - \mu_b)] \end{cases} \tag{8}$$

$$\Longrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \Sigma^{-1}\mu = A_{aa}\mu_a - A_{ab}(x_b-\mu_b) \end{cases} \tag{9}$$

$$\Longrightarrow \begin{cases} \Sigma^{-1} = A_{aa} \\ \mu = \mu_a - A_{aa}^{-1}A_{ab}(x_b-\mu_b) \end{cases} \tag{10}$$

By using Schur complement:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D_{-1}CMBD^{-1} \end{pmatrix},$$

with M = (A - BD$^{-1}C)^{-1}$ (11)

$$\Longrightarrow \begin{cases} A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} \\ A_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-a} \end{cases} \tag{12}$$

As a result:s

$$\begin{cases} \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \\ \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{cases} \Longrightarrow p(x_a|x_b) = \mathcal{N}(x_{a|b}|\mu_{a|b}, \Sigma_{a|b}) \tag{13}$$

(b) The marginal of Gaussian Distribution

The marginal is given by:

$$p(x_a) = \int p(x_a, x_b) dx_2$$

Recall the quadratic form of Gaussian distribution, We need to integrate out x_b by considering the terms involving x_2. Implementing, From the quadratic form

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

We can write:

$$-\frac{1}{2} x_b^T A_{bb} x_b + x_b^T [A_{bb}\mu_b - A_{ba}(x_a - \mu_a)]$$

Let

$$m = A_{bb}\mu_b - A_{ba}(x_a - \mu_a)$$

We got

$$-\frac{1}{2} x_b^T A_{bb} x_b + x_b^T [A_{bb}\mu_b - A_{ba}(x_a - \mu_a)] = -\frac{1}{2} x_b^T A_{bb} x_b + x_a^T m = -\frac{1}{2}(x_b - A_b b^{-1} m)^T A_{bb}(x_b - A_{bb}^{-1} m) + \frac{1}{2} m^T A_{bb}^{-1} m$$

We can integrate over non-normalized 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 is:

$$-\frac{1}{2} x_a^T (A_{aa} - A_{ab} A_{bb}^{-1} A_{ba}) x_a + x_a^T (A_{aa} - A_{ab} Abb^{-1} A_{ba})^{-1} \mu_a + const \tag{14}$$

Similarly, we have

$$\begin{aligned} E[x_a] &= \mu_a \\ cov[x_a] &= \Sigma_{aa} \\ \implies p(x_a) &= \mathcal{N}(x_a | \mu_a, \Sigma_{aa}) \end{aligned}$$

This is an "Overdue Submission". My apology. ☹