Bayes Inference and Gaussian Process

Siqiao Xue

Version 1.1, 2019-08-30.

Abstract

This document introduces Gaussian distribution and Gaussian process from a mathematical perspective. Base on this, the principles and application of Bayesian inference is summarized.

Contents

1 Multivariate Gaussian distribution

1.1 Probability density function

A vector-valued random variable $\boldsymbol{x} \in \mathbb{R}^n$ follows a multivariate Gaussian distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}^n_{++}$ if the probability density function of $\boldsymbol{x}$ is

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

We can write as $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note that $\mathbb{S}^n_{++}$ refers to the space of symmetric positive definite $n \times n$ matrix.

1.2 Marginal and conditional distribution

Consider a random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Suppose that variables in $\boldsymbol{x}$ has been partitioned into two sets $\boldsymbol{x}_A = (x_1, ...x_r)^\top \in \mathbb{R}^r$ and $\boldsymbol{x}_A = (x_{r+1}, ...x_n)^\top \in \mathbb{R}^{n-r}$ (similarly for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$), such that

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}$$

Here $\boldsymbol{\Sigma}_{AB} = \boldsymbol{\Sigma}_{BA}^\top$ since $\boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top] = \boldsymbol{\Sigma}^\top$.

In many situations it is convenient to work with the inverse inverse of the covariance matrix $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$. and according to the same partitioning of the vector $\boldsymbol{x}$ the partitioned form of $\boldsymbol{\Lambda}$ follows

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{AA} & \boldsymbol{\Lambda}_{AB} \\ \boldsymbol{\Lambda}_{BA} & \boldsymbol{\Lambda}_{BB} \end{pmatrix}$$

Lemma 1.1. *The partitioning of the vector $\boldsymbol{x}$ leads to an equivalent expression of the quadratic form of $\boldsymbol{x}$*

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_A-\boldsymbol{\mu}_A) + (\boldsymbol{x}_B-\boldsymbol{b})^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_B-\boldsymbol{b})$$

where $\boldsymbol{b} = \boldsymbol{\mu}_B + \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}(\boldsymbol{x}_A-\boldsymbol{\mu}_A)$.

Proof. By substituting the partitioned form of $\boldsymbol{x}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$, one obtains

$$\begin{aligned}
& (\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) & (1.1)\\
= & \left(\begin{bmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix}\right)^\top \begin{pmatrix} \boldsymbol{\Lambda}_{AA} & \boldsymbol{\Lambda}_{AB} \\ \boldsymbol{\Lambda}_{BA} & \boldsymbol{\Lambda}_{BB} \end{pmatrix} \left(\begin{bmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix}\right) & \\
= & (\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_A-\boldsymbol{\mu}_A) + (\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AB}(\boldsymbol{x}_B-\boldsymbol{\mu}_B) & \\
& +(\boldsymbol{x}_B-\boldsymbol{\mu}_B)^\top \boldsymbol{\Lambda}_{BA}(\boldsymbol{x}_A-\boldsymbol{\mu}_A) + (\boldsymbol{x}_B-\boldsymbol{\mu}_B)^\top \boldsymbol{\Lambda}_{BB}(\boldsymbol{x}_B-\boldsymbol{\mu}_B) & \\
= & (\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_A-\boldsymbol{\mu}_A) + 2(\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AB}(\boldsymbol{x}_B-\boldsymbol{\mu}_B) + (\boldsymbol{x}_B-\boldsymbol{\mu}_B)^\top \boldsymbol{\Lambda}_{BB}(\boldsymbol{x}_B-\boldsymbol{\mu}_B) & (1.2)
\end{aligned}$$

From Lemma(3.1) we construct the relations between $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$

$$\begin{aligned}
\boldsymbol{\Lambda}_{AA} &= \boldsymbol{\Sigma}_{AA}^{-1} + \boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}(\boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1} \\
\boldsymbol{\Lambda}_{BB} &= \left(\boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}^\top\right)^{-1} \\
\boldsymbol{\Lambda}_{AB} &= -\boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}(\boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}^{-1})^{-1}
\end{aligned}$$

and plug in into the Equation(1.2). Using elementary calculation and the techniques in Lemma(3.2) we will finally arrives at the conclusion. □

Lemma 1.2. *The joint density function equals to the product of two Gaussian densities, i.e.*

$$p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = p(\boldsymbol{x}_A;\boldsymbol{\mu}_A,\boldsymbol{\Sigma}_{AA})p(\boldsymbol{x}_B;\boldsymbol{b},\boldsymbol{V}_{AA})$$

where $\boldsymbol{b} = \boldsymbol{\mu}_B + \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}(\boldsymbol{x}_A-\boldsymbol{\mu}_A)$ *and* $\boldsymbol{V} = \boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}$

Proof. For notation simplicity, set $Q(\boldsymbol{x}_A,\boldsymbol{x}_B) = (\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, $Q_1(\boldsymbol{x}_A) = (\boldsymbol{x}_A-\boldsymbol{\mu}_A)^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_A-\boldsymbol{\mu}_A)$, $Q_2(\boldsymbol{x}_A,\boldsymbol{x}_B) = (\boldsymbol{x}_B-\boldsymbol{b})^\top \boldsymbol{\Lambda}_{AA}(\boldsymbol{x}_B-\boldsymbol{b})$.

By the definition of the joint density

$$\begin{aligned}
p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|}\exp\left(-\frac{1}{2}Q(\boldsymbol{x}_A,\boldsymbol{x}_B)\right) \\
&= \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{AA}|^{1/2}|\boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{AB}^\top \boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}|^{1/2}}\exp\left(-\frac{1}{2}Q(\boldsymbol{x}_A,\boldsymbol{x}_B)\right) \quad \text{(Apply Theorem(3.2) to } |\boldsymbol{\Sigma}|) \\
&= \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_{AA}|^{1/2}}\exp\left(Q_1(\boldsymbol{x}_A)\right)\frac{1}{(2\pi)^{q/2}|\boldsymbol{V}|^{1/2}}\exp\left(Q_2(\boldsymbol{x}_A,\boldsymbol{x}_B)\right) \quad \text{(Apply Lemma(1.1) to } Q(\boldsymbol{x}_A,\boldsymbol{x}_B)) \\
&= p(\boldsymbol{x}_A;\boldsymbol{\mu}_A,\boldsymbol{\Sigma}_{AA})p(\boldsymbol{x}_B;\boldsymbol{b},\boldsymbol{V}_{AA})
\end{aligned}$$

□

Theorem 1.1. *The marginal densities*

$$\begin{aligned}
p(\boldsymbol{x}_A) &= \int p(\boldsymbol{x}_A,\boldsymbol{x}_B;\boldsymbol{\mu},\boldsymbol{\Sigma})\mathrm{d}\boldsymbol{x}_B \\
p(\boldsymbol{x}_B) &= \int p(\boldsymbol{x}_A,\boldsymbol{x}_B;\boldsymbol{\mu},\boldsymbol{\Sigma})\mathrm{d}\boldsymbol{x}_A
\end{aligned}$$

are Gaussian:

$$\begin{aligned}
\boldsymbol{x}_A &\sim \mathcal{N}(\boldsymbol{\mu}_A,\boldsymbol{\Sigma}_{AA}) \\
\boldsymbol{x}_B &\sim \mathcal{N}(\boldsymbol{\mu}_B,\boldsymbol{\Sigma}_{BB})
\end{aligned}$$

Proof. The marginal density of $\boldsymbol{x}_A$ can be simplied to

$$\begin{aligned} p(\boldsymbol{x}_A) &= \int p(\boldsymbol{x}_A, \boldsymbol{x}_B; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{x}_B \\ &= \int p(\boldsymbol{x}_A; \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}) p(\boldsymbol{x}_B; \boldsymbol{b}, \boldsymbol{V}_{AA}) \mathrm{d}\boldsymbol{x}_B \quad \text{(Because of Lemma1.2)} \\ &= p(\boldsymbol{x}_A; \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}) \quad \text{(Because } p(\boldsymbol{x}_A; \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}) \text{ is independent of } \boldsymbol{x}_B) \end{aligned}$$

The density function $p(\boldsymbol{x}_A)$ indicates the distribution of $\boldsymbol{x}_A$. By symmetric techniques we arrive at the same distribution for $\boldsymbol{x}_B$.

□

Theorem 1.2. *The conditional densities*

$$p(\boldsymbol{x}_A | \boldsymbol{x}_B), p(\boldsymbol{x}_B | \boldsymbol{x}_A)$$

are Gaussian:

$$\begin{aligned} \boldsymbol{x}_A | \boldsymbol{x}_B &\sim \mathcal{N}\left(\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{\boldsymbol{AB}} \boldsymbol{\Sigma}_{\boldsymbol{BB}}^{-1} (\boldsymbol{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_{\boldsymbol{AA}} - \boldsymbol{\Sigma}_{\boldsymbol{AB}} \boldsymbol{\Sigma}_{\boldsymbol{AB}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{BA}}\right) \\ \boldsymbol{x}_B | \boldsymbol{x}_A &\sim \mathcal{N}\left(\boldsymbol{\mu}_B + \boldsymbol{\Sigma}_{\boldsymbol{BA}} \boldsymbol{\Sigma}_{\boldsymbol{AA}}^{-1} (\boldsymbol{x}_A - \boldsymbol{\mu}_A), \boldsymbol{\Sigma}_{\boldsymbol{BB}} - \boldsymbol{\Sigma}_{\boldsymbol{BA}} \boldsymbol{\Sigma}_{\boldsymbol{AA}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{AB}}\right) \end{aligned}$$

Proof. The conditional density

$$\begin{aligned} p(\boldsymbol{x}_B | \boldsymbol{x}_A) &= \frac{p(\boldsymbol{x}_A, \boldsymbol{x}_B)}{p(\boldsymbol{x}_A)} \\ &= \frac{p(\boldsymbol{x}_A; \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}) p(\boldsymbol{x}_B; \boldsymbol{b}, \boldsymbol{V}_{AA})}{p(\boldsymbol{x}_A; \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}))} \quad \text{(Because of Lemma(1.2) and Theorem(1.1)))} \\ &= p(\boldsymbol{x}_B; \boldsymbol{b}, \boldsymbol{V}_{AA}) \end{aligned}$$

which coincides with the second Gaussian density in the theorem. Same procedure applies to $p(\boldsymbol{x}_A | \boldsymbol{x}_B)$ completes the proof. □

2 Gaussian process

Definition 2.1. *A Gaussian process $X(t), t \geq 0$ is a stochastic process with the property that for every set of times $0 \leq t_1 \leq t_2 \leq ... \leq t_n$, the set of random variables*

$$X(t_1), X(t_2), ..., X(t_n)$$

is jointly normally distributed.

If X_t is a Gaussian process then its distribution is determined by its mean function

$$m(t) = \mathbb{E} X_t$$

and its covariance function

$$\rho(s, t) = \mathbb{E}[(X(s) - m(s)) \dot{(} X(t) - m(t))]$$

3 Appendix

3.1 Theorems on linear algebra

3.1.1 Inverse and determinant of partitioned symmetric matrix

Lemma 3.1. *Assume $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are four matrix, then*

$$(\boldsymbol{A} + \boldsymbol{CBD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \boldsymbol{C} (\boldsymbol{B}^{-1} + \boldsymbol{D} \boldsymbol{A}^{-1} \boldsymbol{C})^{-1} \boldsymbol{D} \boldsymbol{A}^{-1}$$

Proof.

$$
\begin{aligned}
& (\boldsymbol{A}+\boldsymbol{CBD})\left(\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}\right)\\
= & (\boldsymbol{A}+\boldsymbol{CBD})\boldsymbol{A}^{-1}-(\boldsymbol{A}+\boldsymbol{CBD})\boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}\\
= & \boldsymbol{I}+\boldsymbol{CBDA}^{-1}-(\boldsymbol{C}+\boldsymbol{CBDA}^{-1}\boldsymbol{C})(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}\\
= & \boldsymbol{I}+\boldsymbol{CBDA}^{-1}-\boldsymbol{CB}(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}\\
= & \boldsymbol{I}+\boldsymbol{CBDA}^{-1}-\boldsymbol{CBDA}^{-1}\\
= & \boldsymbol{I}
\end{aligned}
$$

□

Theorem 3.1. *Assume a symmetric $n \times n$ matrix A is divided into four blocks*

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{\top} & \boldsymbol{A}_{22}\end{bmatrix}$$

where $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}$ is $p \times p, q \times q$ matrix and $p+q=n$. Then we have the inverse matrix of $\boldsymbol{B}^{-1}=\boldsymbol{A}$ can also be partitioned into

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{12}^{\top} & \boldsymbol{B}_{22}\end{bmatrix}$$

where

$$
\begin{aligned}
&\boldsymbol{B}_{11}=\left(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}\right)^{-1}=\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\\
&\boldsymbol{B}_{22}=\left(\boldsymbol{A}_{22}-\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}^{\top}\right)^{-1}=\boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{22}^{-1}\\
&\boldsymbol{B}_{12}^{\top}=-\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top})^{-1}\\
&\boldsymbol{B}_{12}=-\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}^{-1})^{-1}
\end{aligned}
$$

Proof. By definition of the inverse matrix, we obtains

$$\boldsymbol{I}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{\top} & \boldsymbol{A}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{12}^{\top} & \boldsymbol{B}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11}\boldsymbol{B}_{11}+\boldsymbol{A}_{12}\boldsymbol{B}_{12}^{\top} & \boldsymbol{A}_{11}\boldsymbol{B}_{12}+\boldsymbol{A}_{12}\boldsymbol{B}_{22}\\ \boldsymbol{A}_{12}^{\top}\boldsymbol{B}_{11}+\boldsymbol{A}_{22}\boldsymbol{B}_{12}^{\top} & \boldsymbol{A}_{12}^{\top}\boldsymbol{B}_{12}+\boldsymbol{A}_{22}\boldsymbol{B}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}_p & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_q\end{bmatrix}$$

By equaling corresponding blocks we have

$$\boldsymbol{A}_{11}\boldsymbol{B}_{11}+\boldsymbol{A}_{12}\boldsymbol{B}_{12}^{\top}=\boldsymbol{I}_p \tag{3.1}$$

$$\boldsymbol{A}_{11}\boldsymbol{B}_{12}+\boldsymbol{A}_{12}\boldsymbol{B}_{22}=\boldsymbol{0} \tag{3.2}$$

$$\boldsymbol{A}_{12}^{\top}\boldsymbol{B}_{11}+\boldsymbol{A}_{22}\boldsymbol{B}_{12}^{\top}=\boldsymbol{0} \tag{3.3}$$

$$\boldsymbol{A}_{12}^{\top}\boldsymbol{B}_{12}+\boldsymbol{A}_{22}\boldsymbol{B}_{22}=\boldsymbol{I}_q \tag{3.4}$$

From Equation(3.2) we find $\boldsymbol{B}_{12}=-\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}$ and plug it into Equation(3.1) to get

$$\boldsymbol{B}_{11}=\left(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}\right)^{-1} \tag{3.5}$$

Applying Lemma(3.1) to Equation(3.5) we obtains an equivalent expression

$$\boldsymbol{B}_{11}=\left(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}\right)^{-1}=\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}$$

The same procedure applies to $\boldsymbol{B}_{22}, \boldsymbol{B}_{12}^{\top}, \boldsymbol{B}_{12}$ to achieve the conclusion.

□

Theorem 3.2. *Assume matrix $\boldsymbol{A}$ is partitioned as*

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$$

, *then its determinant*

$$|\boldsymbol{A}| = |\boldsymbol{A_{22}}| \left|\boldsymbol{A}_{11} - \boldsymbol{A}^{-1}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}\right| = |\boldsymbol{A_{11}}| \left|\boldsymbol{A}_{22} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right|$$

Proof. Note that

$$\begin{aligned} \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_{11} & 0 \\ \boldsymbol{A}_{12}^{\top} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \\ 0 & \boldsymbol{A}_{22} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_{12} \\ 0 & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top} & 0 \\ \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & I \end{bmatrix} \end{aligned}$$

By elementary matrix operations, we find

$$\left|\begin{bmatrix} \boldsymbol{A}_{11} & 0 \\ \boldsymbol{A}_{12}^{\top} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \\ 0 & \boldsymbol{A}_{22} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix}\right| = \left|\boldsymbol{A}_{11}(\boldsymbol{A}_{22} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})\right| = |\boldsymbol{A_{11}}| \left|\boldsymbol{A}_{22} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right|$$

With the same reasoning, the following can be obtained

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_{12} \\ 0 & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} - \boldsymbol{A}_{12}^{\top}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top} & 0 \\ \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & I \end{bmatrix} = |\boldsymbol{A_{22}}| \left|\boldsymbol{A}_{11} - \boldsymbol{A}^{-1}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^{\top}\right|$$

Then the theorem is proved.

□

Lemma 3.2. *Assume* $\boldsymbol{A}$ *is a symmetric matrix, then the following equation holds for any vectors* $\boldsymbol{u}, \boldsymbol{v}$*.*

$$\boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^{\top}\boldsymbol{A}\boldsymbol{v} = (\boldsymbol{v} - \boldsymbol{u})^{\top}A(\boldsymbol{v} - \boldsymbol{u})$$

Proof.

$$\begin{aligned} & \boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^{\top}\boldsymbol{A}\boldsymbol{v} \\ = \; & \boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{u} - \boldsymbol{u}^{\top}\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^{\top}\boldsymbol{A}\boldsymbol{v} - \boldsymbol{u}^{\top}A\boldsymbol{v} \\ = \; & \boldsymbol{u}^{\top}\boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v}) + (\boldsymbol{u} - \boldsymbol{v})^{\top}\boldsymbol{A}\boldsymbol{v} \\ = \; & (\boldsymbol{u} - \boldsymbol{v})^{\top}\boldsymbol{A}\boldsymbol{u} + (\boldsymbol{u} - \boldsymbol{v})^{\top}\boldsymbol{A}\boldsymbol{v} \quad \left(\text{Because } \boldsymbol{A} \text{ is symmetric, } \boldsymbol{u}^{\top}\boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v}) = (\boldsymbol{u} - \boldsymbol{v})^{\top}\boldsymbol{A}\boldsymbol{u}\right) \\ = \; & (\boldsymbol{v} - \boldsymbol{u})^{\top}\boldsymbol{A}(\boldsymbol{v} - \boldsymbol{u}) \end{aligned}$$

□

References

[1] Chuong B. Do, Gaussian Process, lecture note of Standard University, 2007.

[2] G. Pages, Introduction to Numerical Probability, lecture note of École Polytechnique, 2011.

[3] D. Revuz and M. Yor. Continuous martingales and Brownian motion, 2nd edition. *Springer*, 1994.

[4] R. Wang. Marginal and conditional distributions of multivariate normal distribution, *http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html*, 2006.