8 Gaussian Process

In order to understand the Gaussian Process, we need to become conversant with manipulating multivariate normal distributions with density

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

where $|\Sigma|$ is the determinant of a positive definite covariance matrix Σ.

8.1 Schur Complement

Let M be an invertible $(n+m)\times(n+m)$ matrix in a block form

$$M = \begin{array}{cc} & \begin{array}{cc} n & m \end{array} \\ & \left(\begin{array}{cc} A & B \\ C & D \end{array}\right) \begin{array}{c} n \\ m \end{array} \end{array}.$$

Because M is invertible, there exists a unique solution to the linear equation

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{8.1}$$

If D is invertible, then we can use Gaussian elimination to solve for x and y:

$$y = D^{-1}\beta - D^{-1}Cx \;\Rightarrow\; (M/D)x = \alpha - BD^{-1}\beta,$$

where $M/D := A - BD^{-1}C$ is called the Schur complement of D in M; from the block determinant formula (Theorem A.3),

$$\det(M) = \det(D)\det(M/D),$$

we see that (M/D) is also invertible when M and D are both invertible. The solution to (8.1) is then

$$\begin{aligned} x &= (M/D)^{-1}(\alpha - BD^{-1}\beta) \\ y &= D^{-1}\left[\beta - C(M/D)^{-1}(\alpha - BD^{-1}\beta)\right]. \end{aligned}$$

Combining these two equations into a matrix form, we get

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

where

$$\boxed{\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1} \end{pmatrix}}. \tag{8.2}$$

Similar, if A and its Schur complement $M/A := D - CA^{-1}B$ are invertible, then

$$\boxed{\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{pmatrix}}. \tag{8.3}$$

If A, M/A, D, and M/D are all invertible, then we can combine (8.2) and (8.3) to get

$$\boxed{\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (M/D)^{-1} & -A^{-1}B(M/A)^{-1} \\ -D^{-1}C(M/D)^{-1} & (M/A)^{-1} \end{pmatrix}}, \tag{8.4}$$

which reproduces the formula (A.6).

Theorem 8.1. *If* $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ *is a positive definite matrix, then* A*,* M/A*,* D*, and* M/D *are all invertible and* (8.4) *holds.*

Proof. We immediately see that A and D are invertible, because they are principal submatrices of M and, thus, must be positive definite. The block determinant formula (Theorem A.3) then implies that M/A and M/D have positive determinants and are thus invertible. In fact, we can further show that M/A and M/D are also positive definite, as follows: rewrite (8.2) as

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} \begin{pmatrix} (M/D)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix},$$

which implies that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} M/D & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}.$$

(Aside: Note that this last equation is valid whenever D is invertible and does not require that M/D is also invertible.) When M is symmetric, which is the case for any positive definite matrix, D is symmetric and $C = B^t$. Hence, the last equation becomes

$$\begin{pmatrix} A & B \\ B^t & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} M/D & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}^t.$$

Since $\begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}$ is invertible, we now see that $\begin{pmatrix} M/D & 0 \\ 0 & D \end{pmatrix}$ has to be positive definite. Thus, M/D has to be positive definite. A similar argument shows that

$$\begin{pmatrix} A & B \\ B^t & D \end{pmatrix} = \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}^t \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix},$$

which implies the M/A has to be positive definite. □

8.2 Marginal and Conditional Distributions of Multivariate Normal

Define the precision matrix Λ to be the inverse of the covariance matrix Σ. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a multivariate normal random vector with mean $\boldsymbol{\mu}$, and $\boldsymbol{X_a} = (X_1, \ldots, X_k)$ and $\boldsymbol{X_b} = (X_{k+1}, \ldots, X_n)$, for $1 \leq k < n$. Then, we can decompose Λ in a block form as

$$\Lambda = \Sigma^{-1} = \begin{pmatrix} \boldsymbol{\Lambda_{aa}} & \boldsymbol{\Lambda_{ab}} \\ \boldsymbol{\Lambda_{ba}} & \boldsymbol{\Lambda_{bb}} \end{pmatrix}.$$

Using (8.2), we see that

$$\boldsymbol{\Lambda_{aa}} = (\Sigma/\boldsymbol{\Sigma_{bb}})^{-1} \Rightarrow (\boldsymbol{\Lambda_{aa}})^{-1} = \Sigma/\boldsymbol{\Sigma_{bb}}.$$

By exchanging the role of Σ and Λ, we also get

$$(\boldsymbol{\Sigma_{aa}})^{-1} = \Lambda/\boldsymbol{\Lambda_{bb}}. \tag{8.5}$$

Using this last equation, we can integrate out $\boldsymbol{X_b}$ to prove

Theorem 8.2 (Marginal Distribution)**.** *The marginal distribution of $\boldsymbol{X_a}$ is multivariate normal with mean $\boldsymbol{\mu_a}$ and covariance matrix $Cov(\boldsymbol{X_a}, \boldsymbol{X_a})$.*

Proof. Problem Set. □

Similarly, treating $\boldsymbol{X_b}$ as a constant vector shows that

Theorem 8.3 (Conditional Distribution)**.** *The conditional distribution of $\boldsymbol{X_a}$ given $\boldsymbol{X_b} = \boldsymbol{x_b}$ is multivariate normal with mean*

$$E[\boldsymbol{X_a}|\boldsymbol{X_b} = \boldsymbol{x_b}] = \boldsymbol{\mu_a} + \boldsymbol{\Sigma_{ab}\Sigma_{bb}^{-1}}(\boldsymbol{x_b} - \boldsymbol{\mu_b})$$

and variance

$$Var[\boldsymbol{X_a}|\boldsymbol{x_b}] = \boldsymbol{\Sigma_{aa}} - \boldsymbol{\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}} = \Sigma/\boldsymbol{\Sigma_{bb}} = \boldsymbol{\Lambda_{aa}^{-1}}.$$

Proof. See the proof of Theorem A.17. □