Gaussian Process

Marshal Sinaga - marshal.sinaga@aalto.fi

Last update: March 11, 2024

This note aims to cover some materials on the Gaussian process. The primary references are Gaussian Process for Machine Learning by C. E. Rasmussen and CS-E4895 by Arno Solin.

1 Multivariate Normal Distribution

1.1 Linear transformation theorem for the multivariate normal distribution

Let x follow a multivariate normal distribution:

$$x \sim \mathcal{N}(\mu, \Sigma) \tag{1}$$

Then, any affine transformation of x is also multivariate normally distributed:

$$y = Ax + b \sim \mathcal{N}(A\mu + b, A\Sigma A^\top) \tag{2}$$

Proof:

The moment-generating function of random vector x is

$$M_x(t) = \mathbb{E}[\exp(t^T x)] \tag{3}$$

and therefore, the moment-generating function of the random vector y is given by

$$\begin{aligned} M_y(t) &= \mathbb{E}\left[\exp(t^T(Ax + b))\right] \\ &= \mathbb{E}[\exp(t^\top Ax)\exp(t^\top b)] \\ &= \exp(t^\top b)\mathbb{E}[\exp(t^\top Ax)] \\ &= \exp(t^\top b)M_x(A^\top t) \end{aligned} \tag{4}$$

The moment-generating function of the multivariate normal distribution is

$$M_x(t) = \exp(t^\top \mu + \frac{1}{2}t^\top \Sigma t) \tag{5}$$

and therefore, the moment-generating function of random vector y becomes

$$M_y(t) = \exp(t^\top(A\mu + b) + \frac{1}{2}t^\top A\Sigma A^\top t) \tag{6}$$

Since the moment-generating function and the probability density function of a random variable are equivalent, this demonstrates that y follows a multivariate normal distribution with mean $A\mu + b$ and covariance $A\Sigma A^\top$.

1.2 Marginal distribution of the multivariate normal distribution

Let x follow a multivariate normal distribution:

$$x \sim \mathcal{N}(\mu, \Sigma) \tag{7}$$

Then, the marginal distribution of any subset vector x_s is also a multivariate normal distribution

$$x_s \sim \mathcal{N}(\mu_s, \Sigma_s) \tag{8}$$

where μ_s drops the irrelevant variables (the ones not in the subset, i.e., marginalized out) from the mean vector μ and Σ_s drops the corresponding rows and columns from the covariance matrix Σ.

Proof: Define an $m \times n$ subset matrix S such that $s_{ij} = 1$, if the $j-$th element in x_s corresponds to the $i-$th element in x, and $s_{ij} = 0$ otherwise. Then,

$$x_s = Sx \tag{9}$$

and we can apply the linear transformation theorem to give

$$x_s \sim \mathcal{N}(S\mu, S\Sigma S^\top) \tag{10}$$

Finally, we see that $S\mu = \mu_s$ and $S\Sigma S^\top = \Sigma_s$

1.3 Conditional distribution of the multivariate normal distribution

Let x follow a multivariate normal distribution

$$x \sim \mathcal{N}(\mu, \Sigma) \tag{11}$$

Then, the conditional distribution of any subset vector x_1, given the complement vector x_2, is also a multivariate normal distribution

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned} \tag{12}$$

with block-wise mean and covariance defined as:

$$\begin{aligned} \mu &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \end{aligned} \tag{13}$$

Proof: Without loss of generality, we assume that in parallel to 13,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{14}$$

where $x_1 \in \mathbb{R}^{n_1 \times 1}$, $x_2 \in \mathbb{R}^{n_2 \times 1}$, and $x \in \mathbb{R}^{n \times 1}$. The joint distribution of x_1 and x_2 is

$$x \sim \mathcal{N}(\mu, \Sigma) \tag{15}$$

Moreover, the marginal distribution of x_2 follows from 11 and 13 as

$$x_2 \sim \mathcal{N}(\mu_2, \Sigma_{22}) \tag{16}$$

According to conditional probability, it holds that

$$\begin{aligned} p(x_1|x_2) &= \frac{p(x_1, x_2)}{p(x_2)} \\ &= \frac{\mathcal{N}(\mu, \Sigma)}{\mathcal{N}(\mu_2, \Sigma_{22})} \end{aligned} \tag{17}$$

Using the probability density of multivariate-normal, this becomes

$$\begin{aligned} p(x_1|x_2) &= \frac{1/\sqrt{(2\pi)^n |\Sigma|} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)}{1/\sqrt{(2\pi)^{n_2}|\Sigma_{22}|} \exp\left(-\frac{1}{2}(x-\mu_2)^\top \Sigma_{22}^{-1}(x-\mu_2)\right)} \\ &= 1/\sqrt{(2\pi)^{n-n_2}} \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu) + \frac{1}{2}(x-\mu_2)^\top \Sigma_{22}^{-1}(x-\mu_2)\right) \end{aligned} \tag{18}$$

Writing the inverse Σ as

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{19}$$

and applying 13 to 18, we obtain:

$$p(x_1|x_2) = 1/\sqrt{(2\pi)^{n-n_2}} \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \exp(-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^\top \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right) \tag{20}$$

$$+ \frac{1}{2}(x-\mu_2)^\top \Sigma_{22}^{-1}(x-\mu_2))$$

Multiplying within 20, we have

$$\begin{aligned} p(x_1|x_2) = 1/\sqrt{(2\pi)^{n-n_2}} \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \exp(-\frac{1}{2}((x_1-\mu_1)^\top \Sigma^{11}(x_1-\mu_1) + 2(x_1-\mu_1)\Sigma^{12}(x_2-\mu_2) \\ + (x_2-\mu_2)^\top \Sigma^{22}(x_2-\mu_2)) + \frac{1}{2}(x-\mu_2)^\top \Sigma_{22}^{-1}(x-\mu_2)) \end{aligned} \tag{21}$$

where we have used the fact that $\Sigma^{12} = \Sigma^{21^\top}$, because Σ^{-1} is symmetric. The inverse of a block matrix is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{bmatrix} \tag{22}$$

Thus, the inverse of Σ^{-1} in 19 is

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{bmatrix} \tag{23}$$

Plugging this into 20, we have

$$\begin{aligned} p(x_1|x_2) =& \frac{1}{\sqrt{(2\pi)^{n-n_2}}} \cdot \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \cdot \\ & \exp\left[-\frac{1}{2}\left((x_1 - \mu_1)^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} (x_1 - \mu_1) \; - \right.\right. \\ & \qquad 2(x_1 - \mu_1)^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) + \\ & \qquad (x_2 - \mu_2)^{\mathrm{T}} \left[\Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}\right](x_2 - \mu_2)\Big) \\ & \quad \left. + \frac{1}{2}\left((x_2 - \mu_2)^{\mathrm{T}}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right] \; . \end{aligned} \tag{24}$$

Eliminating some terms, we have

$$\begin{aligned} p(x_1|x_2) =& \frac{1}{\sqrt{(2\pi)^{n-n_2}}} \cdot \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \cdot \\ & \exp\left[-\frac{1}{2}\left((x_1 - \mu_1)^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} (x_1 - \mu_1) \; - \right.\right. \\ & \qquad 2(x_1 - \mu_1)^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) + \\ & \qquad \left.\left.(x_2 - \mu_2)^{\mathrm{T}}\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right] \; . \end{aligned} \tag{25}$$

Rearranging the terms, we have

$$\begin{aligned} p(x_1|x_2) =& \frac{1}{\sqrt{(2\pi)^{n-n_2}}} \cdot \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \cdot \exp\left[-\frac{1}{2} \cdot \right. \\ & \left.\left[(x_1 - \mu_1) - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right]^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \left[(x_1 - \mu_1) - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right]\right] \\ =& \frac{1}{\sqrt{(2\pi)^{n-n_2}}} \cdot \sqrt{\frac{|\Sigma_{22}|}{|\Sigma|}} \cdot \exp\left[-\frac{1}{2} \cdot \right. \\ & \left.\left[x_1 - \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right]^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \left[x_1 - \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right]\right] \end{aligned} \tag{26}$$

where we used the fact that $\Sigma_{21} = \Sigma_{12}^{\top}$. The determinant of a block matrix is

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot |A - BD^{-1}C| \; , \tag{27}$$

such that we have for Σ that

$$\begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} = |\Sigma_{22}| \cdot |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}| \tag{28}$$

with this and $n - n_2 = n_1$, we finally arrive at

$$\begin{aligned} p(x_1|x_2) = & \frac{1}{\sqrt{(2\pi)^{n_1}|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|}} \cdot \exp\left[-\frac{1}{2}\cdot\right. \\ & \left.\left[x_1 - \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right]^{\mathrm{T}} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \left[x_1 - \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)\right]\right] \end{aligned} \tag{29}$$

which is the pdf of a multivariate normal distribution

$$p(x_1|x_2) = \mathcal{N}(x_1; \mu_{1|2}, \Sigma_{1|2}) \tag{30}$$

with mean $\mu_{1|2}$ and covariance $\Sigma_{1|2}$ given by 12.