L03: Gaussians II

Gonzalo Ferrer

2 February 2023

1 Conditioning a joint Gaussian PDF

$$p(x_a, x_b) = \alpha \exp \underbrace{\left\{ -\frac{1}{2} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix}^\intercal \Sigma^{-1} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} \right\}}_{\Delta}$$

Problem: $\quad p(x_a|x_b), \quad$ where $x_a \in \mathcal{R}^n, x_b \in \mathcal{R}^m$

$\Sigma = \begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix}, \quad \Sigma_{ab} = \Sigma_{ba}^\intercal$ (Σ symmetric)

Information matrix $\Lambda = \Sigma^{-1}, \quad \Lambda = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$

Expand the exponent Δ:

$$\begin{aligned} \Delta = &-\frac{1}{2}(x_a - \mu_a)^\intercal \Lambda_a (x_a - \mu_a) - \frac{1}{2}(x_a - \mu_a)^\intercal \Lambda_{ab}(x_b - \mu_b) \\ &-\frac{1}{2}(x_b - \mu_b)^\intercal \Lambda_{ba}(x_a - \mu_a) - \frac{1}{2}(x_b - \mu_b)^\intercal \Lambda_b (x_b - \mu_b) \end{aligned} \tag{1}$$

Solution: Completing the square

Intuition $\longrightarrow$ we want an exponent to only depend on x_a since x_b is conditioned ("given")

$$\Delta = -\frac{1}{2} x_a^\intercal \Sigma_{a|b}^{-1} x_a + x_a^\intercal \Sigma_{a|b}^{-1} m - \frac{1}{2} m^\intercal \Sigma_{a|b}^{-1} m + const \tag{2}$$

Q: What happens to x_b and constant terms?

2nd order term: $x_a^\intercal \Sigma_{a|b}^{-1} x_a$,

$$\Sigma_{a|b}^{-1} = \Lambda_a \tag{3}$$

1st order term: $x_a^\intercal \underbrace{(\Lambda_a \mu_a - \Lambda_{ab}(x_b - \mu_b))}_{\Sigma_{a|b}^{-1}\mu_{a|b}} \implies$

$$\mu_{a|b} = \Sigma_{a|b}(\underbrace{\Lambda_a}_{\text{use (3)}} \mu_a - \Lambda_{ab}(x_b - \mu_b)) = \mu_a - \Sigma_{a|b}\Lambda_{ab}(x_b - \mu_b) \tag{4}$$

We'll use the following matrix equality:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}+D^{-1}CMBD^{-1} \end{bmatrix} \tag{5}$$

Where $M = (A - BD^{-1}C)^{-1}$, $\quad$ (M^{-1} Schur complement)

Let $\begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$, then

$$\Lambda_a = (\overbrace{\Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba}}^{\Sigma_{a|b}})^{-1} \tag{6}$$

$$\Lambda_{ab} = -\Lambda_a\Sigma_{ab}\Sigma_b^{-1} \tag{7}$$

$$\begin{aligned} \implies \mu_{a|b} &= \mu_a - \Sigma_{a|b}\overbrace{(-\underbrace{\Lambda_a}_{\text{use (3)}}\Sigma_{ab}\Sigma_b^{-1})}^{\Lambda_{ab}}(x_b - \mu_b) \\ &= \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b - \mu_b) \end{aligned} \tag{8}$$

$$p(x_a|x_b) = \mathcal{N}(x_a; \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b - \mu_b), \Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba})$$

2 Marginalizing a joint Gaussian PDF

Problem: $\quad p(x_a) = \int p(x_a, x_b)dx_b$

Solution: $\quad$ Same as before, we will expand the exponent and complete the square, now twice

$$\int e^{\Delta}dx_b = e^{\Delta(x_a)}\underbrace{\int e^{\Delta(x_b)}dx_b}_{\eta} * \underbrace{e^{\Delta(const)}}_{\eta'}$$

$$\begin{aligned} \Delta &= -\frac{1}{2}(x_a-\mu_a)^\intercal\Lambda_a(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^\intercal\Lambda_{ab}(x_b-\mu_b) \\ &\quad -\frac{1}{2}(x_b-\mu_b)^\intercal\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^\intercal\Lambda_b(x_b-\mu_b) \\ &= \underbrace{-\frac{1}{2}x_b^\intercal\Lambda_b x_b + x_b^\intercal \overbrace{(\Lambda_b\mu_b - \Lambda_{ba}(x_a-\mu_a))}^{\Lambda_b m \implies m = \mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)} - \frac{1}{2}m^\intercal\Lambda_b m}_{\Delta(x_b)} \\ &\quad \underbrace{+\frac{1}{2}m^\intercal\Lambda_b m - \frac{1}{2}x_a^\intercal\Lambda_a x_a + x_a^\intercal(\Lambda_a\mu_a + \Lambda_{ab}\mu_b) + const}_{\Delta(x_a)} \end{aligned} \tag{9}$$

$$\begin{aligned}
\frac{1}{2}m^{\intercal}\Lambda_b m &= \frac{1}{2}(\mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a))^{\intercal}\Lambda_b(\mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)) \\
&= \frac{1}{2}x_a^{\intercal}\Lambda_{ab}\Lambda_b^{-1}\Lambda_b\Lambda_b^{-1}\Lambda_{ba}x_a - x_a^{\intercal}\Lambda_{ab}\Lambda_b^{-1}\Lambda_b(\mu_b + \Lambda_b^{-1}\Lambda_{ba}\mu_a) + const \\
&= \frac{1}{2}x_a^{\intercal}\Lambda_{ab}\Lambda_b^{-1}\Lambda_{ba}x_a - x_a^{\intercal}\Lambda_{ab}(\mu_b + \Lambda_b^{-1}\Lambda_{ba}\mu_a) + const
\end{aligned} \tag{10}$$

$$\begin{aligned}
\Delta(x_a) &= -\frac{1}{2}x_a^{\intercal}\Lambda_a x_a + \frac{1}{2}x_a^{\intercal}\Lambda_{ab}\Lambda_b^{-1}\Lambda_{ba}x_a + x_a^{\intercal}(\Lambda_a\mu_a + \Lambda_{ab}\mu_b) \\
&\quad - x_a^{\intercal}\Lambda_{ab}(\mu_b + \Lambda_b^{-1}\Lambda_{ba}\mu_a) + const \\
&= -\frac{1}{2}x_a^{\intercal}\underbrace{\overset{\text{Schur complement!}}{(\Lambda_a - \Lambda_{ab}\Lambda_b^{-1}\Lambda_{ba})}}_{\Sigma_a}x_a + x_a^{\intercal}(\Lambda_a - \Lambda_{ab}\Lambda_b^{-1}\Lambda_{ba})\mu_a + const
\end{aligned} \tag{11}$$

$$p(x_a) = \int p(x_a, x_b)dx_b = \mathcal{N}(x_a; \mu_a, \Sigma_a)$$

- Marginalizing a Gaussian is as simple as selecting the submatrix inside Σ and the corresponding μ!
- Gaussians are their self conjugate priors