Gaussian Processes with Uncertain Inputs: Predictions and Derivatives

Carl Edward Rasmussen

November 24th, 2014

We want to model data with E output coordinates, and use seperate combinations of linear models and GPs to make predictions, $a = 1, \dots, E$:

$$f_a(x^*) \;=\; f_a^* \;\sim\; \mathcal{N}\big(\theta_a^\top x^* + k_a(x^*, \mathbf{x})\beta_a,\; k_a(x^*, x^*) - k_a(x^*, \mathbf{x})(K_a + \Sigma_\varepsilon^a)^{-1} k_a(\mathbf{x}, x^*)\big),$$

where the E squared exponential covariance functions are

$$k_a(x, x') \;=\; s_a^2 q(x, x', \Lambda_a, 0), \;\; \text{where} \;\; a = 1, \dots, E, \tag{1}$$

and s_a^2 are the signal variances and Λ_a is a diagonal matrix of squared length scales for GP number a. The noise variances are Σ_ε^a. The inputs are $\mathbf{x}$ and the outputs y_a and we define $\beta_a = (K_a + \Sigma_\varepsilon^a)^{-1}(y_a - \theta_a^\top \mathbf{x})$.

Predictions at uncertain inputs

Consider making predictions from $a = 1, \dots, E$ GPs at $\mathbf{x}^*$ with specificati on

$$p(\mathbf{x}^* | \mathbf{m}, \Sigma) \;\sim\; \mathcal{N}(\mathbf{m}, \Sigma). \tag{2}$$

We have the following expressions for the predictive mean, variances and input output covariances

$$\begin{aligned}
\mathbb{E}[\mathbf{f}^* | \mathbf{m}, \Sigma] \;&=\; \int \big(s_a^2 \beta_a^\top q(x_i, \mathbf{x}^*, \Lambda_a, 0) + \theta_a^\top \mathbf{x}^*\big)\mathcal{N}(\mathbf{x}^* | \mathbf{m}, \Sigma) d\ bfx^* \;=\; s_a^2 \beta_a^\top q^a + \theta_a^\top \mathbf{m}, & (3)\\
\mathbb{C}[x^*, f_a^* | \mathbf{m}, \Sigma] \;&=\; \int (x^* - \mathbf{m})\big(s_a^2 \beta_a^\top q(\mathbf{x}, x^*, \Lambda_a, 0) + \theta_a^\top x^*\big)\mathcal{N}(x^* | \mathbf{m}, \Sigma) dx^* \\
&=\; s_a^2 \Sigma(\Lambda_a + \Sigma)^{-1}(\mathbf{x} - \mathbf{m})\beta_a q^a + \Sigma\theta_a \;=\; \Sigma C_a + \Sigma\theta_a, & (4)\\
\mathbb{V}[f_a^* | \mathbf{m}, \Sigma] \;&=\; \mathbb{V}[\mathbb{E}[f_a^* | x^*] | \mathbf{m}, \Sigma] + \mathbb{E}[\mathbb{V}[f_a^* | x^*] | \mathbf{m}, \Sigma] \\
&=\; \mathbb{V}[s_a^2 \beta_a^\top q(\mathbf{x}, x^*, \Lambda_a, 0) + \theta_a^\top x^*] + \delta_{ab}\mathbb{E}[s_a^2 - k_a(x^*, \mathbf{x})(K_a + \Sigma_\varepsilon^a)^{-1} k_a(\mathbf{x}, x^*)] & (5)\\
&=\; s_a^2 s_b^2 \big[\beta_a^\top (Q^{ab} - q^a q^{b\top})\beta_b + \delta_{ab}\big(s_a^{-2} - \mathrm{tr}((K_a + \Sigma_\varepsilon^a)^{-1} Q^{aa})\big)\big] + C_a^\top \Sigma\theta_b + \theta_a^\top \Sigma C_b + \theta_a^\top \Sigma\theta_b, \\
\text{where} \quad q_i^a \;&=\; q(x_i, \mathbf{m}, \Lambda_a, \Sigma), \;\; \text{and} \;\; Q_{ij}^{ab} \;=\; Q\big(x_i, x_j, \Lambda_a, \Lambda_b, 0, \mathbf{m}, \Sigma\big).
\end{aligned}$$

In the above we've made use of the following two functions

$$
\begin{aligned}
q(x, x', \Lambda, V) &\triangleq |\Lambda^{-1}V + I|^{-1/2} \exp\big(-\tfrac{1}{2}(x-x')[\Lambda + V]^{-1}(x-x')\big), \\
Q(x, x', \Lambda_a, \Lambda_b, V, \mu, \Sigma) &\triangleq c_1 \exp\big(-\tfrac{1}{2}(x-x')^\top[\Lambda_a + \Lambda_b + 2V]^{-1}(x-x')\big) \\
&\quad \times \exp\big(-\tfrac{1}{2}(z-\mu)^\top\big[\big((\Lambda_a+V)^{-1} + (\Lambda_b+V)^{-1}\big)^{-1} + \Sigma\big]^{-1}(z-\mu)\big), \\
&= c_2\, q(x, \mu, \Lambda_a, V)\, q(\mu, x'\Lambda_b, V) \\
&\quad \times \exp\big(\tfrac{1}{2}\mathbf{r}^\top\big[(\Lambda_a+V)^{-1} + (\Lambda_b+V)^{-1} + \Sigma^{-1}\big]^{-1}\mathbf{r}\big), \\
&\text{where} \begin{cases} z \;=\; (\Lambda_b+V)(\Lambda_a+\Lambda_b+2V)^{-1}x + (\Lambda_a+V)(\Lambda_a+\Lambda_b+2V)^{-1}x' \\ \mathbf{r} \;=\; (\Lambda_a+V)^{-1}(x-\mu) + (\Lambda_b+V)^{-1}(x'-\mu) \\ c_1 \;=\; \big|(\Lambda_a+V)(\Lambda_b+V) + (\Lambda_a+\Lambda_b+2V)\Sigma\big|^{-1/2}\big|\Lambda_a\Lambda_b\big|^{1/2} \\ c_2 \;=\; \big|\big((\Lambda_a+V)^{-1} + (\Lambda_b+V)^{-1}\big)\Sigma + I\big|^{-1/2}, \end{cases}
\end{aligned} \tag{6}
$$