GP Prediction with Hierarchical Uncertain Inputs

Carl Edward Rasmussen and Rowan McAllister

December 23nd, 2014

The two functions

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

have the following Gaussian integrals

$$
\begin{aligned}
\int q(x, t, \Lambda, V)\mathcal{N}(t|\mu, \Sigma)dt &= q(x, \mu, \Lambda, \Sigma + V), \\
\int q(x, t, \Lambda_a, V)\, q(t, x', \Lambda_b, V)\, \mathcal{N}(t|\mu, \Sigma)dt &= Q(x, x', \Lambda_a, \Lambda_b, V, \mu, \Sigma), \\
\int Q(x, x', \Lambda_a, \Lambda_b, 0, \mu, V)\mathcal{N}(\mu|\mathbf{m}, \Sigma)d\mu &= Q(x, x', \Lambda_a, \Lambda_b, 0, \mathbf{m}, \Sigma + V).
\end{aligned} \tag{2}
$$

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

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

where the E squared exponential covariance functions are

$$
k_a(x, x') \;=\; s_a^2 q(x, x', \Lambda_a, 0), \quad \text{where} \quad a = 1, \ldots, E, \tag{3}
$$

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 $a = 1, \ldots, E$ GPs at $\mathbf{x}^*$ with specification

$$p(\mathbf{x}^*|\mathbf{m}, \Sigma) \;\sim\; \mathcal{N}(\mathbf{m}, \Sigma). \tag{4}$$

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

$$\begin{aligned}
\mathbb{E}[\mathbf{f}^*|\mathbf{m},\Sigma] \;&=\; \int \big(s_a^2 \beta_a^\top q(x_i, \mathbf{x}^*, \Lambda_a, 0) + \theta_a^\top \mathbf{x}^*\big)\mathcal{N}(\mathbf{x}^*|\mathbf{m},\Sigma)d\mathbf{x}^* \;=\; s_a^2 \beta_a^\top q^a + \theta_a^\top \mathbf{m}, && (5)\\
\mathbb{C}[x^*, f_a^*|\mathbf{m},\Sigma] \;&=\; \int (x^* - \mathbf{m})\big(s_a^2 \beta_a^\top q(\mathbf{x}, x^*, \Lambda_a, 0) + \theta_a^\top x^*\big)\mathcal{N}(x^*|\mathbf{m},\Sigma)dx^* \\
&=\; s_a^2 \Sigma(\Lambda_a + \Sigma)^{-1}(\mathbf{x} - \mathbf{m})\beta_a q^a + \Sigma\theta_a \;=\; \Sigma C_a + \Sigma\theta_a, && (6)\\
\mathbb{V}[f_a^*|\mathbf{m},\Sigma] \;&=\; \mathbb{V}[\mathbb{E}[f_a^*|x^*]|\mathbf{m},\Sigma] + \mathbb{E}[\mathbb{V}[f_a^*|x^*]|\mathbf{m},\Sigma] \\
&=\; \mathbb{V}[s_a^2 \beta_a^\top q(\mathbf{x}, x^*, \Lambda_a, 0) + \theta_a^\top x^*] + \delta_{ab}\mathbb{E}[s_a^2 - k_a(x^*, \mathbf{x})(K_a + \Sigma_\varepsilon^a)^{-1}k_a(\mathbf{x}, x^*)] && (7)\\
&=\; s_a^2 s_b^2 \big[\beta_a^\top (Q^{ab} - q^a q^{b\top})\beta_b + \delta_{ab}\big(s_a^{-2} - \operatorname{tr}((K_a + \Sigma_\varepsilon^a)^{-1}Q^{aa})\big)\big] + C_a^\top \Sigma\theta_b + \theta_a^\top \Sigma C_b + \theta_a^\top \Sigma\theta_b,\\
\text{where}\quad q_i^a \;&=\; q(x_i, \mathbf{m}, \Lambda_a, \Sigma), \quad \text{and} \quad Q_{ij}^{ab} \;=\; Q\big(x_i, x_j, \Lambda_a, \Lambda_b, 0, \mathbf{m}, \Sigma\big).
\end{aligned}$$

Predictions at hierarchical uncertain inputs

Consider making predictions from $a = 1, \ldots, E$ GPs at $\mathbf{x}^*$ with *hierarchical* specification

$$p(\mathbf{x}^*|\mu) \;\sim\; \mathcal{N}(\mu, V), \quad \text{and} \quad \mu \;\sim\; \mathcal{N}(\mathbf{m}, \Sigma), \tag{8}$$

or equivalently the joint

$$p\Big(\begin{bmatrix} \mathbf{x}^* \\ \mu \end{bmatrix}\Big) \;\sim\; \mathcal{N}\Big(\begin{bmatrix} \mathbf{m} \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} \Sigma + V & \Sigma \\ \Sigma & \Sigma \end{bmatrix}\Big). \tag{9}$$

We're interested in the following quantities

$$\mathbb{E}[\mathbb{E}[f(x^*|\mu,V)]], \quad \mathbb{C}[\mu, \mathbb{E}[f(x^*|\mu,V)]], \quad \mathbb{V}[\mathbb{E}[f(x^*|\mu,V)]], \quad \mathbb{E}[\mathbb{C}[x^*, f(x^*|\mu,V)]] \;\text{ and }\; \mathbb{E}[\mathbb{V}[f(x^*|\mu,V)]]. \tag{10}$$

For the *mean of the mean* we have

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[f(x^*|\mu,V)]] \;&=\; \int \mathbb{E}[f(x^*|\mu,V)]\mathcal{N}(\mu|\mathbf{m},\Sigma)d\mu \\
&=\; s_a^2 \beta_a^\top \int q(\mathbf{x}, \mu, \Lambda_a, V)\mathcal{N}(\mu|\mathbf{m},\Sigma)d\mu + \theta_a^\top \mathbf{m} \;=\; s_a^2 \beta_a^\top q(\mathbf{x}, \mathbf{m}, \Lambda_a, \Sigma + V) + \theta_a^\top \mathbf{m}.
\end{aligned} \tag{11}$$

For the *covariance of the mean* we have

$$\begin{aligned}
\mathbb{C}[\mu, \mathbb{E}[f(x^*|\mu,V)]] \;&=\; \int (\mu - \mathbf{m})\mathbb{E}[f(x^*|\mu,V)]\mathcal{N}(\mu|\mathbf{m},\Sigma)d\mu \\
&=\; \int (\mu - \mathbf{m})\big(s_a^2 \beta_a^\top q(x_i, \mu, \Lambda_a, V) + \theta_a^\top \mu\big)\mathcal{N}(\mu|\mathbf{m},\Sigma)d\mu \\
&=\; s_a^2 \Sigma(\Lambda_a + \Sigma + V)^{-1}(\mathbf{x} - \mathbf{m})\beta_a q(\mathbf{x}, \mathbf{m}, \Lambda_a, \Sigma + V) + \Sigma\theta_a \;=\; \Sigma\hat{C}_a + \Sigma\theta_a.
\end{aligned} \tag{12}$$

For the *variance of the mean* we have

$$
\begin{aligned}
\mathbb{V}[\mathbb{E}[f(x^*|\mu,V)]] &= \int \mathbb{E}[f(x^*|\mu,V)]^2 \mathcal{N}(\mu|\mathbf{m},\Sigma)\,d\mu - \mathbb{E}[\mathbb{E}[f(x^*|\mu,V)]]^2 \\
&= s_a^2 s_b^2 \beta_a^\top (\hat{Q}^{ab} - q^a q^{b\top})\beta_b + \hat{C}_a^\top \Sigma\theta_b + \theta_a^\top \Sigma \hat{C}_b + \theta_a^\top \Sigma\theta_b, \\
\text{where} \quad q_i^a &= q(x_i,\mathbf{m},\Lambda_a,\Sigma+V), \quad \text{and} \quad \hat{Q}_{ij}^{ab} \;=\; Q(x_i,x_j,\Lambda_a,\Lambda_b,V,\mathbf{m},\Sigma).
\end{aligned}
\tag{13}
$$

For the *mean of the covariance* we have

$$
\begin{aligned}
\mathbb{E}[\mathbb{C}[x^*,f(x^*|\mu,V)]] &= \int \mathbb{C}[x^*,f(x^*|\mu,V)]\mathcal{N}(\mu|\mathbf{m},\Sigma)\,d\mu, \\
&= s_a^2 V(\Lambda_a+V)^{-1}\int(\mathbf{x}-\mu)\beta_a^\top q(\mathbf{x},\mu,\Lambda_a,V)\mathcal{N}(\mu|\mathbf{m},\Sigma)\,d\mu + V\theta_a, \\
&= s_a^2 V(\Lambda_a+\Sigma+V)^{-1}(\mathbf{x}-\mathbf{m})\beta_a^\top q(\mathbf{x},\mathbf{m},\Lambda_a,\Sigma+V) + V\theta_a \;=\; V\hat{C}_a + V\theta_a.
\end{aligned}
\tag{14}
$$

Finally, for the *mean of the variance* we have

$$
\begin{aligned}
\mathbb{E}[\mathbb{V}[f(x^*|\mu,V)] &= \int \mathbb{V}[f(x^*|\mu,V)]\mathcal{N}(\mu|\mathbf{m},\Sigma)\,d\mu \qquad (15) \\
&= s_a^2 s_b^2 \big[\beta_a^\top(\tilde{Q}^{ab} - \hat{Q}^{ab})\beta_b + \delta_{ab}\big(s_a^{-2} - \mathrm{tr}((K_a+\Sigma_\varepsilon^a)^{-1}\tilde{Q}^{aa})\big)\big] + \hat{C}_a^\top V\theta_b + \theta_a^\top V\hat{C}_b + \theta_a^\top V\theta_b, \\
\text{where} \quad \tilde{Q}_{ij}^{ab} &= Q(x_i,x_j,\Lambda_a,\Lambda_b,0,\mathbf{m},\Sigma+V), \quad \text{and} \quad \hat{Q}_{ij}^{ab} \;=\; Q(x_i,x_j,\Lambda_a,\Lambda_b,V,\mathbf{m},\Sigma).
\end{aligned}
$$

Note, that for the special case $V = 0$, eq. (5) is equal to eq. (11), eq. (7) is equal to the sum of eq. (13) and eq. (15) and eq. (6) is equal to eq. (12).

Derivatives

For symmetric Λ and V and Σ:

$$
\begin{aligned}
\frac{\partial \ln q(x,x',\Lambda,V)}{\partial x} &= -(\Lambda+V)^{-1}(x-x') = -(\Lambda^{-1}V+I)^{-1}\Lambda^{-1}(x-x') \\
\frac{\partial \ln q(x,x',\Lambda,V)}{\partial x'} &= (\Lambda+V)^{-1}(x-x') \\
\frac{\partial \ln q(x,x',\Lambda,V)}{\partial V} &= -\frac{1}{2}(\Lambda+V)^{-1} + \frac{1}{2}(\Lambda+V)^{-1}(x-x')(x-x')^\top(\Lambda+V)^{-1}
\end{aligned}
\tag{16}
$$

Let $L = (\Lambda_a + V)^{-1} + (\Lambda_b + V)^{-1}$, $R = \Sigma L + I$, $Y = R^{-1}\Sigma = \left[L + \Sigma^{-1}\right]^{-1}$, $T : X \to XX^\top$:

$$
\begin{aligned}
\partial Q(x, x', \Lambda_a, \Lambda_b, V, \mu, \Sigma) \;&=\; Q \circ \partial\Big(\ln c_2 + \ln q(x, \mu, \Lambda_a, V) + \ln q(\mu, x'\Lambda_b, V) + \frac{1}{2} y^\top Y y \Big) \\
\frac{1}{2}\frac{\partial\, y^\top Y y}{\partial \mu} &= y^\top Y \frac{\partial y}{\partial \mu} = -y^\top Y L \\
\frac{\partial \ln c_2}{\partial \Sigma} &= -\frac{1}{2}\frac{\partial \ln |L\Sigma + I|}{\partial \Sigma} = -\frac{1}{2} L^\top (L\Sigma + I)^{-\top} = -\frac{1}{2} L R^{-1} \\
\frac{\partial\, y^\top Y y}{\partial \Sigma} &= \Sigma^{-\top} Y^\top y y^\top Y^\top \Sigma^{-\top} = T(R^{-\top} y) \\
\frac{\partial \ln c_2}{\partial V} &= -\frac{1}{2}\frac{\partial \ln |L\Sigma + I|}{\partial V} = -\frac{1}{2}\frac{\partial \ln |\sum_i \left[(\Lambda_i + V)^{-1}\right]\Sigma + I|}{\partial V} \\
&= \frac{1}{2}\sum_i \Big[(\Lambda_i + V)^{-\top} \Big(\sum_j \left[(\Lambda_j + V)^{-1}\right]\Sigma + I \Big)^{-\top} \Sigma^\top (\Lambda_i + V)^{-\top} \Big] \\
&= \frac{1}{2}\sum_i \Big[(\Lambda_i + V)^{-1} Y (\Lambda_i + V)^{-1} \Big] \\
\frac{\partial\, y^\top Y y}{\partial V} &= y^\top \frac{\partial\, Y}{\partial V} y + \frac{\partial y^\top}{\partial V} Y y + y^\top Y \frac{\partial y}{\partial V} \quad = \quad \sum_i \Big[(\Lambda_i + V)^{-1} Y^\top y y^\top Y^\top (\Lambda_i + V)^{-1} \Big] \\
&\quad - \sum_i \Big[(\Lambda_i + V)^{-1} (x_{n_i} - \mu)(Y y)^\top (\Lambda_i + V)^{-1} \Big] - \sum_i \Big[(\Lambda_i + V)^{-1} (y^\top Y)^\top (x_{n_i} - \mu)^\top (\Lambda_i + V)^{-1} \Big] \\
&= \sum_i \Big[T\Big((\Lambda_i + V)^{-1} (Y y - (x_{n_i} - \mu)) \Big) - T\Big((\Lambda_i + V)^{-1} (x_{n_i} - \mu) \Big) \Big]
\end{aligned}
\tag{17}
$$