Derivação do método de integração que aproxima pelo quadrado

Danilo de Freitas Naiff

Idéia (*Sampling for Inference in Probabilistic Models with Fast Bayesian Quadrature*): Temos uma integral $\mathbb{E}[l(X)] = \int l(x)p(x)dx$, que queremos aproximar, *sabendo que* l *é uma verossimilhança*. Assumimos $p(x)$ Gaussiana $\mathcal{N}(x;\nu,\Lambda)$ (em geral multivariada). Aproximamos então $\tilde{l}(x) = \sqrt{2l(x)}$ (no paper original tem um parametro α próximo de zero, mas tiramos por enquanto) por um processo Gaussiano. Logo $l(x) = \frac{1}{2}\tilde{l}(x)^2$. Temos então, que dado $x_\mathcal{D}$ pontos onde $\tilde{l}$ é medido, temos que

$$\tilde{m}_\mathcal{D}(x) = \mathbf{k}^T(x,x_\mathcal{D})K_\mathcal{D}^{-1}\tilde{l}(x_\mathcal{D}) \tag{1}$$

. Fazendo uma linearização (um pouco roubada), temos que

$$l(x) \approx \frac{1}{2}\tilde{m}_\mathcal{D}(x)^2, \tag{2}$$

De onde podemos integrar. Talvez seja mais fácil pensar na aproximação acima como uma regressão da verossimilhança, que preserva:

- Positividade
- Extrapolação para zero fora da região de amostragem.

Adaptamos essa idéia para a aproximação da marginalização de um Processo Gaussiano, inspirado fortemente em *Gaussian Processes for Prediction.*

Temos em geral:

$$p(y|\mathcal{D},I) = \frac{1}{Z}\int p(y|\phi,\mathcal{D},I)p(\phi|\mathcal{D},I)p(\phi|I)d\phi \tag{3}$$

A partir de agora tiramos a dependência de I por conveniência. Sabemos que, dado ϕ, $p(y|\phi,\mathcal{D}) = f_{N(y|\mu_\phi,\Sigma_\phi)}(x)$. Então, dado ϕ_S amostras dos hiperparâmetros, consideramos $l(\phi) = p(y|\phi,\mathcal{D},I)p(\phi|\mathcal{D},I)$. Assim (e usamos aqui fortemente o *Matrix Cookbook*):

$$\tilde{l}(\phi) = \sqrt{2}(4\pi)^{n/4}|\Sigma_\phi|^{1/4}f_{N(\mu_\phi,2\Sigma_\phi)}(y)\sqrt{a(\phi)} \tag{4}$$

Onde $a(\phi) = p(\phi|\mathcal{D}, I)$. Então, aproximamos (3) por:

$$\begin{aligned}
p(y|\mathcal{D}, \phi_s) \approx & \frac{1}{Z} \int \frac{1}{2} \tilde{m}_{\phi_s}(\phi) p(\phi) d\phi \\
= & \frac{1}{2Z} \int \tilde{l}(\phi_s)^T K_{\phi_s}^{-1} \mathbf{k}(\phi, \phi^s) \mathbf{k}(\phi, \phi^s)^T K_{\phi_s}^{-1} \tilde{l}(\phi_s) p(\phi) d\phi \\
= & \frac{(4\pi)^{n/2}}{Z} \sum_{i,j} \Big(M_{i,j} |\Sigma_{(\phi_i)}|^{1/4} |\Sigma_{(\phi_j)}|^{1/4} \sqrt{a(\phi_i)} \sqrt{a(\phi_j)} \\
& f_{N(\mu_{\phi_i}, 2\Sigma_{\phi_i})}(y) f_{N(\mu_{\phi_j}, 2\Sigma_{\phi_j})}(y) \Big)
\end{aligned} \tag{5}$$

Onde:

$$\begin{aligned}
M &= K_{\phi_s}^{-1} W K_{\phi_s}^{-1} \\
W_{i,j} &= \int k(\phi, \phi_i) k(\phi, \phi_j) p(\phi) d\phi
\end{aligned} \tag{6}$$

Mas temos que:

$$\begin{aligned}
& f_{N(\mu_{\phi_i}, 2\Sigma_{\phi_i})}(y) f_{N(\mu_{\phi_j}, 2\Sigma_{\phi_j})}(y) = C(\phi_i, \phi_j) f_{N(\mu_{i,j}, \Sigma_{i,j})}(y) \\
& C(\phi_i, \phi_j) = \frac{1}{(4\pi)^{n/2} |\Sigma_{\phi_i} + \Sigma_{\phi_j}|} \exp\Big(- \frac{1}{4} (\mu_{\phi_i} - \mu_{\phi_j})^T (\Sigma_{\phi_i} + \Sigma_{\phi_j})^{-1} (\mu_{\phi_i} - \mu_{\phi_j}) \Big) \\
& \mu_{i,j} = (\Sigma_{\phi_i}^{-1} + \Sigma_{\phi_j}^{-1})^{-1} (\Sigma_{\phi_i}^{-1} \mu_{\phi_j} + \Sigma_{\phi_j}^{-1} \mu_{\phi_i}) \\
& \Sigma_{i,j} = 2(\Sigma_{\phi_i}^{-1} + \Sigma_{\phi_j}^{-1})^{-1}
\end{aligned} \tag{7}$$

Juntando os termos constantes em uma nova constante de normalização (também chamada Z), temos então que:

$$p(y|\mathcal{D}, \phi_s) \approx \frac{1}{Z} \sum_{i,j} M_{i,j} |\Sigma_{(\phi_i)}|^{1/4} |\Sigma_{(\phi_j)}|^{1/4} \sqrt{a(\phi_i) a(\phi_j)} C(\phi_i, \phi_j) f_{N(\mu_{i,j}, \Sigma_{i,j})}(y) \tag{8}$$

Para calcular W, assumimos que $k(\phi_1, \phi_2)$ é um *squared-exponential*, ou seja, $k(\phi_1, \phi_2) = h^2 \exp(-\frac{1}{2} \sum_{k=1}^d (\phi_1^{(k)} - \phi_2^{(k)})^2 / l_k^2)$, e que $p(\phi) = f_{\mathcal{N}(\nu, diag(\lambda^2))}(\phi) = \prod_{k=1}^d f_{N(\nu_k, \lambda_k^2)}(\phi)$, temos que:

$$W_{i,j} = h^2 \prod_{k=1}^d \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\lambda_k^2}} \exp\Big(- \frac{1}{2} \Big(\frac{(\phi^{(k)} - \phi_j^{(k)})^2}{l_k^2} + \frac{(\phi^{(k)} - \phi_j^{(k)})^2}{l_k^2} + + \frac{(\phi^{(k)} - \nu_k)^2}{\lambda_k^2} \Big) \Big) \tag{9}$$

Simplificando a integral acima, temos que, completando quadrado,

$$\begin{aligned}
& W_{i,j} = h^2 \prod_{k=1}^d \Big(\frac{V_k^2}{\lambda_k^2} \Big)^{1/2} \exp\Big(- \frac{1}{2} C_k \Big) \\
& V_k^2 = (2l_k^{-2} + \lambda_k^{-2})^{-1} \\
& C_k = l_k^{-2} ((\phi_i^{(k)})^2 + (\phi_j^{(k)})^2) + \lambda_k^{-2} \nu_k^2 - (2l_k^{-2} + \lambda_k^{-2})^{-1} (l_k^{-2} ((\phi_i^{(k)}) + (\phi_j^{(k)})) + \lambda_k^{-2} \nu_k)^2
\end{aligned} \tag{10}$$