Gaussian Process object which handles Angles

Rowan McAllister

June 26, 2015

Titsias method: "Variational Learning of Inducing Variables in Sparse Gaussian Processes"

Using Eq.9 from Titsias paper, plus extra terms V, C and U:

$$\begin{aligned}
\mathrm{nlml} &\triangleq -\log[\mathcal{N}(y;0,\sigma^2 I_n + Q_{nn})] + \tfrac{1}{2\sigma^2}\mathrm{Tr}(K_{nn} - Q_{nn}) \\
&= \tfrac{1}{2}\big(n\log[2\pi] + \log[\det[\sigma^2 I_n + Q_{nn}]] + y^\top(\sigma^2 I_n + Q_{nn})^{-1}y + \sigma^{-2}\mathrm{Tr}(K_{nn} - Q_{nn})\big) \\
Q_{nn} &\triangleq K_{nm}K_{mm}^{-1}K_{mn} \\
&= V^\top V \\
V &= L^{-\top}K_{mn} = \mathrm{chol}(K_{mm})^\top \backslash K_{mn} \\
K_{mm} &= L^\top L \\
C &\triangleq \mathrm{chol}(\sigma^2 I_m + VV^\top) \\
U &\triangleq (\sigma^2 I_m + VV^\top)^{-\frac{\top}{2}} = C^\top \backslash V
\end{aligned} \tag{1}$$

Using matrix inversion lemma:

$$\begin{aligned}
(\sigma^2 I_n + Q_{nn})^{-1} &= (\sigma^2 I_n + K_{nm}K_{mm}^{-1}K_{mn})^{-1} \\
&= \sigma^{-2}I_n - \sigma^{-2}K_{nm}(K_{mm} + \sigma^{-2}K_{mn}K_{nm})^{-1}K_{mn}\sigma^{-2} \\
&= \sigma^{-2}\big(I_n - K_{nm}(\underbrace{\sigma^2 L^\top L + K_{mn}K_{nm}}_{A})^{-1}K_{mn}\big) \\
&= \sigma^{-2}\big(I_n - K_{nm}L^{-1}(\sigma^2 I_m + L^{-\top}K_{mn}K_{nm}L^{-1})^{-1}L^{-\top}K_{mn}\big) \\
&= \sigma^{-2}\big(I_n - V^\top(\sigma^2 I_m + VV^\top)^{-1}V\big) \\
&= \sigma^{-2}\big(I_n - U^\top U\big)
\end{aligned} \tag{2}$$

For the determinant we make use of identity:

$$\det(aI_n \pm CA^{-1}B) = a^{n-m}\det(aA \pm BC)/\det(A) \quad \text{where } A \in \mathcal{R}^{m\times m}, B \in \mathcal{R}^{m\times n}, C \in \mathcal{R}^{n\times m}, a \in \mathcal{R} \tag{3}$$

and thus have:

$$\begin{aligned}
\det[\sigma^2 I + Q_{nn}] &= \det[\sigma^2 I + K_{nm}K_{mm}^{-1}K_{mn}] \\
&= \sigma^{2(n-m)}\det[\sigma^2 K_{mm} + K_{mn}K_{nm}]/\det[K_{mm}] \\
&= \sigma^{2(n-m)}\det[\sigma^2 I_m + VV^\top] \\
&= \sigma^{2(n-m)}(\prod \mathrm{diag}(C))^2 \\
\log[\det[\sigma^2 I + Q_{nn}]] &= (n-m)\log[\sigma^2] + 2\sum \log[\mathrm{diag}(C)]
\end{aligned} \tag{4}$$

For the trace:

$$\begin{aligned}\mathsf{Tr}(K_{nn} - Q_{nn}) &= \mathsf{Tr}(K_{nn}) - \mathsf{Tr}(V^\top V) \\ &= n\sigma^2_{\text{signal}} - \sum_{ij} (V \odot V)_{ij}\end{aligned} \tag{5}$$

Derivatives

We use derivatives tricks from Ed Snelson's PhD thesis "Flexible and efficient Gaussian process models for machine learning", appendix C:

$$\begin{aligned}Q_{nn} &\triangleq K_{nm}K_{mm}^{-1}K_{mn} \\ q_{nn} &\triangleq \sigma^2 I_n + Q_{nn} \\ A &= \sigma^2 K_{mm} + K_{mn}K_{nm} \\ \dot{A} &= \sigma^2 \dot{K}_{mm} + \dot{K}_{mn}K_{nm} + K_{mn}\dot{K}_{nm} \\ B &= K_{mn}^\top A^{-1} \\ \frac{\partial}{\partial x_{mf}} nlml &= \tfrac{1}{2}\Big(\frac{\partial}{\partial x_{mf}} \underbrace{\log[\det[q_{nn}]]}_{\mathcal{L}_1} + \frac{\partial}{\partial x_{mf}} \underbrace{y^\top q_{nn}^{-1} y}_{\mathcal{L}_2} + \sigma^{-2} \frac{\partial}{\partial x_{mf}} \underbrace{\mathsf{Tr}(-Q_{nn})}_{\mathcal{L}_3}\Big)\end{aligned} \tag{6}$$

$$\begin{aligned}\mathcal{L}_1 &= \log[\det[q_{nn}]] \\ &= \log[\det[A]] - \log[\det[K_{mm}]] + (N - M)\log \sigma^2 \\ \dot{\mathcal{L}_1} &= \mathsf{Tr}(A^{-\frac{\top}{2}} \dot{A} A^{-\frac{1}{2}}) - \mathsf{Tr}(K_{mm}^{-\frac{\top}{2}} \dot{K}_{mm} K_{mm}^{-\frac{1}{2}}) \quad \text{(see Eq C.8 Ed's thesis)} \\ \\ \mathcal{L}_2 &= y^\top q_{nn}^{-1} y \\ \dot{\mathcal{L}_2} &= \tfrac{1}{\sigma^2} y^\top [K_{mn}^\top A^{-\frac{1}{2}} (A^{-\frac{\top}{2}} \dot{A} A^{-\frac{1}{2}} (A^{-\frac{\top}{2}} K_{mn}) - 2A^{-\frac{\top}{2}} \dot{K}_{mn})] y \quad \text{(see Eq C.9 Ed's thesis)} \\ &= \tfrac{1}{\sigma^2} y^\top [B(\dot{A} B^\top - 2\dot{K}_{mn})] y \\ \\ \mathcal{L}_3 &= \mathsf{Tr}(-Q_{nn}) \\ &= -\mathsf{Tr}(K_{nm}K_{mm}^{-1}K_{mn}) \\ \dot{\mathcal{L}_3} &= -\mathsf{Tr}(\dot{Q}_{nn}) \\ &= -\mathsf{Tr}(-K_{nm}K_{mm}^{-1}\dot{K}_{mm}K_{mm}^{-1}K_{mn} + \dot{K}_{nm}K_{mm}^{-1}K_{mn} + K_{nm}K_{mm}^{-1}\dot{K}_{mn});\end{aligned} \tag{7}$$