The Reward Function

Carl Edward Rasmussen

July 17th 2008

The *reward* R of being in state $\mathbf{s}$ is

$$\mathsf{R}(\mathbf{s}) \;=\; \exp\big(-\tfrac{1}{2}(\mathbf{s}-\mathbf{z})^\top \mathsf{T}^{-1}(\mathbf{s}-\mathbf{z})\big),$$

where $\mathbf{z}$ and T are parameters of the reward function, which typically depend on the specific task. We assume that the state is uncertain, with a Gaussian distribution

$$\mathbf{s} \;\sim\; \mathcal{N}(\mathbf{m}, \mathsf{S}),$$

with mean $\mathbf{m}$ and covariance matrix S. We wish to compute the *expected* reward

$$\mu \;=\; \mathbb{E}[\mathsf{R}] \;=\; \langle \mathsf{R}(\mathbf{s})\rangle_{\mathbf{s}\sim\mathcal{N}(\mathbf{m},\mathsf{S})} \;=\; |\mathsf{I}+\mathsf{S}\mathsf{T}^{-1}|^{-1/2}\exp\big(-\tfrac{1}{2}(\mathbf{m}-\mathbf{z})^\top(\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z})\big),$$

and to compute the *variance* of the reward we first define

$$r^2 \;=\; \mathbb{E}[\mathsf{R}^2] \;=\; |\mathsf{I}+2\mathsf{S}\mathsf{T}^{-1}|^{-1/2}\exp\big(-(\mathbf{m}-\mathbf{z})^\top(2\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z})\big),$$

so that

$$\sigma^2 \;=\; \mathbb{V}[\mathsf{R}] \;=\; \mathbb{E}[\mathsf{R}^2]-\mathbb{E}^2[\mathsf{R}] \;=\; r^2-\mu^2.$$

We also need the derivatives of these two quanteties wrt the parameters of the state distribution. For the expected reward we have

$$\frac{d\mu}{d\mathbf{m}} \;=\; -\mu(\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z}), \qquad \frac{d\mu}{d\mathsf{S}} \;=\; \tfrac{1}{2}\mu\big[(\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z})(\mathbf{m}-\mathbf{z})^\top-\mathsf{I}\big](\mathsf{S}+\mathsf{T})^{-1},$$

and for the variance

$$\frac{d\sigma^2}{d\mathbf{m}} \;=\; -2r^2(2\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z})-2\mu\frac{d\mu}{d\mathbf{m}}, \qquad \frac{d\sigma^2}{d\mathsf{S}} \;=\; r^2\big[2(2\mathsf{S}+\mathsf{T})^{-1}(\mathbf{m}-\mathbf{z})(\mathbf{m}-\mathbf{z})^\top-\mathsf{I}\big](2\mathsf{S}+\mathsf{T})^{-1}-2\mu\frac{d\mu}{d\mathsf{S}}.$$