Probability of Improvement Exploration Heuristic for PILCO

Rowan McAllister

July 10, 2015

Let the parameterisation of the previous rollout be r, and the cumulative-cost distribution given the previous rollout be:

$$\mathcal{C}^r \sim \mathcal{N}(\mu_r, \sigma_r^2) \tag{1}$$

We would like to choose a new parameterisation, θ, in such a way that it maximises the *probability of improvement* (P.I.) of the cumulative-cost. For arbitrary θ we have cumulative-cost distribution $\mathcal{C}^\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$. What is the probability $P(\mathcal{C}^\theta < \mathcal{C}^r)$? Let $\Delta\mathcal{C} \doteq \mathcal{C}^\theta - \mathcal{C}^r$. Note:

$$\Delta\mathcal{C} \quad \sim \quad \mathcal{N}(\mu_\theta - \mu_r, \sigma_\theta^2 + \sigma_r^2 - 2c) \tag{2}$$

where c is the covariance between $\mathcal{C}^\theta$ and $\mathcal{C}^r$. Let's assume (approximate) that $c = 0$, to make life simpler. So now the probability of improvement, by changing parameterisation from r to θ is:

$$\begin{aligned}
\text{P.I.} &= P\big(\mathcal{C}^\theta < \mathcal{C}^r\big) && (3)\\
&= P\big(\Delta\mathcal{C} < 0\big) && (4)\\
&= \int_{-\infty}^{0} \mathcal{N}\big(x; \mu_\theta - \mu_r, \sigma_\theta^2 + \sigma_r^2\big)\,dx && (5)\\
&= 1 - \Phi\Big(\underbrace{\frac{\mu_\theta - \mu_r}{\sqrt{\sigma_\theta^2 + \sigma_r^2}}}_{z}\Big) && (6)
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative standard normal function. We wish to maximise Eq (6), so our loss function (to minimise) is the negative P.I.:

$$L \quad = \quad \Phi(z) - 1, \tag{7}$$

with gradients

$$\begin{aligned}
\frac{dL}{d\mu_\theta} &= \frac{1}{\sqrt{\sigma_\theta^2 + \sigma_r^2}}\phi(z), && (8)\\
\frac{dL}{d\sigma_\theta^2} &= -\frac{1}{2(\sigma_\theta^2 + \sigma_r^2)} z\phi(z), && (9)
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal distribution.