Expected 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 optimises the *expected improvement* (E.I.) of the cumulative-cost. Since we care about low costs, 'improvement' means a decrease in 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 expected improvement, by changing parameterisation from r to θ is:

$$\begin{aligned} \text{E.I.} &= \int_{-\infty}^{0} x\,\mathcal{N}\big(x; \mu_\theta - \mu_r, \sigma_\theta^2 + \sigma_r^2\big)\,dx && (3) \\ &= \Phi(-z)(\mu_\theta - \mu_r) - \phi(z)\sqrt{\sigma_\theta^2 + \sigma_r^2} && (4) \end{aligned}$$

where $\phi(\cdot)$ is the standard normal distribution, $\Phi(\cdot)$ its cumulative standard normal function, and $z = \frac{\mu_\theta - \mu_r}{\sqrt{\sigma_\theta^2 + \sigma_r^2}}$. In this case the E.I. is our loss function.

$$L \quad = \quad \Phi(-z)(\mu_\theta - \mu_r) - \phi(z)\sqrt{\sigma_\theta^2 + \sigma_r^2} \tag{5}$$

With gradients:

$$\begin{aligned} \frac{dL}{d\mu_\theta} &= -\frac{\partial z}{\partial \mu_\theta}\phi(z)(\mu_\theta - \mu_r) + \Phi(-z) + \frac{\partial z}{\partial \mu_\theta} z\phi(z)\sqrt{\sigma_\theta^2 + \sigma_r^2} && (6) \\ &= \Phi(-z) && (7) \\ \frac{dL}{d\sigma_\theta^2} &= -\frac{\partial z}{\partial \sigma_\theta^2}\phi(z)(\mu_\theta - \mu_r) + \frac{\partial z}{\partial \sigma_\theta^2} z\phi(z)\sqrt{\sigma_\theta^2 + \sigma_r^2} - \frac{\phi(z)}{2\sqrt{\sigma_\theta^2 + \sigma_r^2}} \\ &= -\frac{\phi(z)}{2\sqrt{\sigma_\theta^2 + \sigma_r^2}} && (8) \end{aligned}$$