Interleaved Quadrature and Optimization for Planning in Continuous State-Action Space

Nishad Gothoskar

LIS

July 27 2018

Value Function in RL: Discrete vs. Continuous

Discrete

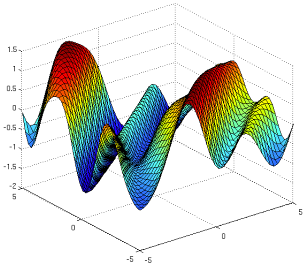
$$V^{\pi}(s) = \max_{a \in A} \Big[\sum_{s' \in S} p_{s'|s,a}(s'|s,a)\big(R(s'|s,a) + \gamma V^{\pi}(s')\big) \Big]$$

Continuous

$$V^{\pi}(s) = \max_{a \in A} \int_{s' \in S} p_{s'|s,a}(s'|s,a)\big(R(s'|s,a) + \gamma^{\Delta t} V^{\pi}(s')\big)\, \mathrm{d}s'$$

In our setting, this involves a maximization in a continuous action space, where the value of an action is an integral over the product of a continuous state space value function and the transition model.

Gaussian Process



We will use a Gaussian Process to model the value function across our state space. With GPs, we define the value at certain support points and allow the model to generalize to other states.

GPs also allow us to understand the distribution of value at a state, which we can use as a representation of our uncertainty about its value.

$$y^*|\mathbf{y} \sim \mathcal{N}(K_* K^{-1}\mathbf{y}, K_{**} - K_* K^{-1} K_*^T)$$