A3Q4: Infinite Continuous MDP with Normal Transitions (Myopic)

Bodhi Nguyen

January 21st, 2021

1 Problem

Consider a continuous-states, continuous-actions, discrete-time, non-terminating MDP with state space as $\mathbb{R}$ and action space as $\mathbb{R}$. When in state $s \in \mathbb{R}$, upon taking action $a \in \mathbb{R}$, one transitions to next state $s' \in \mathbb{R}$ according to a normal distribution $s' \sim N(s, \sigma^2)$ for a fixed variance $\sigma^2 \in \mathbb{R}^+$. The corresponding cost associated with this transition is $e^{as'}$, i.e., the cost depends on the action a and the state s' one transitions to. The problem is to minimize the infinite-horizon Expected Discounted-Sum of Costs (with discount factor $\gamma \leq 1$). For this assignment, solve this problem just for the special case of $\gamma = 0$ (i.e., the myopic case) using elementary calculus. Derive an analytic expression for the optimal action in any state and the corresponding optimal cost.

2 Solution

The optimal value function is

$$V^*(s) = \min_{a \in A}[R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') * V^*(s')] \tag{1}$$

Since $\gamma = 0$, we can eliminate the second part of that equation entirely, and our value function is now to minimize the reward function. The reward function is

$$R(s,a) = E[R_{t+1}|S_t = s, A_t = a] \tag{2}$$

In our problem, this equates to finding the expected value of $e^{as'}$. Since s is normal with a mean of s and variance σ^2, as' is distributed normal with mean of as and variance $a^2\sigma^2$. We now recognize the reward as the moment generating function of the normal distribution, and so

$$R(s,a) = e^{as'} e^{\frac{1}{2}\sigma^2 a^2} \tag{3}$$

which is the MGF of the normal distribution evaluated at $t = 0$. To minimize this, we take the derivative with respect to a:

$$\frac{dR(s,a)}{da} = e^{as'} e^{\frac{1}{2}\sigma^2 a^2}(s' + a\sigma^2) \tag{4}$$

which is equal to zero when $a = \frac{-\sigma^2}{s'}$. Plotting this with variance of 1 and $s' = 1$ shows this critical point is a minimum. The corresponding optimal cost, is thus $e^{-\sigma^2}$.