Advanced ML Assignment 2

May 1, 2023

1 Task 2 - Policy Iteration

1.1 Theory Question 1 - Linear convergence of Policy Iteration

(a)

$$\begin{aligned} BV^{\pi_t}(x) &= \max_a \left[r(x,a) + \gamma \mathbb{E}_{x'|x,a}\left[V^{\pi_t}(x')\right]\right] \\ &\geq r(x,\pi_t(x)) + \gamma \mathbb{E}_{x'|x,\pi_t(x)}\left[V^{\pi_t}(x')\right] \\ &= V^{\pi_t}(x) \end{aligned}$$

(b)

Algorithm 1 Fixed Point Iteration

1: **Initialize** $U_0 = V^{\pi}t$.
2: **for** $i = 1$ to T **do**
3: $U_i = r(x, \pi_{t+1}(x)) + \mathbb{E}x'|x, \pi_{t+1}(x)[U_{i-1}]$
4: **end for**

From fixed point iteration

$$U_1(x) = BV^{\pi_t} \geq V^{\pi_t}(x) = U_0(x)$$

By induction, U_i is monotonically increasing:

$$\begin{aligned} U_i(x) &= r(x,\pi_{t+1}(x)) + E_{x'|x,\pi_{t+1}(x)}[U_{i-1}(x')] \\ &\geq r(x,\pi_{t+1}(x)) + E_{x'|x,\pi_{t+1}(x)}[U_{i-2}(x')] \\ &= U_{i-1}(x) \end{aligned}$$

Therefore

$$V_{\pi_{t+1}} = U_{i\to\infty} \geq U_1 = BV_{\pi_t}$$

(c)

From above, and $V^*(x) \geq V^{\pi}(x)$ for any policy π:

$$||V^{\pi_{t+1}}(x) - V^*||_\infty \leq ||BV^{\pi_t}(x) - V^*||_\infty$$

Because $BV^* = V^*$ and the contractive property of B we proceed by induction:

$$\begin{aligned} ||V^{\pi_{t+1}}(x) - V^*||_\infty &\leq ||BV^{\pi_t}(x) - BV^*||_\infty \\ &\leq ||B(V^{\pi_t}(x) - V^*)||_\infty \\ &\leq \gamma||V^{\pi_t}(x) - V^*||_\infty \\ &\leq \gamma^t||V^{\pi_1}(x) - V^*||_\infty \end{aligned}$$

This expression decays exponentially as t goes to ∞.