ECE 289A - An Introduction to Reinforcement Learning HW#4

Ahmed H. Mahmoud

November, 9th 2017

Q.1

In this question, we regenerated plot in Figure 6.5 in the book. Figure 1 shows the regenerated plots. In order to run the code, please make sure to install MATLAB *Curve Fitting Toolbox* package which is used to smooth the curves.

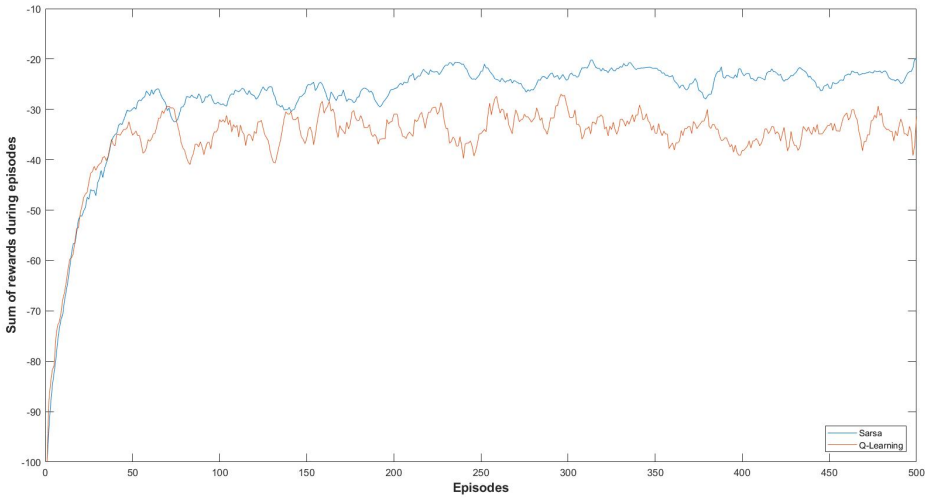

Figure 1

Q.2

Figure 2 shows the regenerate plot from Figure 7.2 in the book.

Q.3

Proving the following inequality for

$$max_s|\mathbb{E}_\pi[G_{t:t+n}|S_t = s] - v_\pi(s)| \leq \gamma^n max_s|V_{t+n-1}(s) - v_\pi(s)|$$

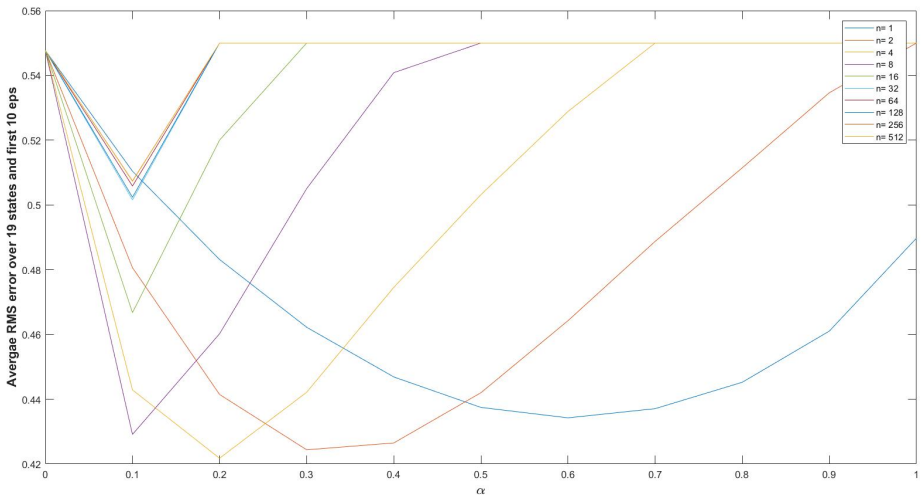

Figure 2: Performance of n-step TD methods as a function of α for various values of n on a 19-state random walk task.

We start by expanding the error estimation

$$max_s|\mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{n-1}R_{t+n} + \gamma^n V_{t+n-1}(S_{t+n})|S_t = s] - v_\pi(s)|$$

In the limit, this quantity should always be less than $\gamma^n max_s|V_{t+n-1}(s) - v_\pi(s)|$ since for any stat s, the accumulated discounted rewards will provide a correction for the estimated $v_\pi(s)$.