Reinforcement Learnig: Homework 1

Raphaël Avalos

November 7, 2018

1 Dynamic Programming

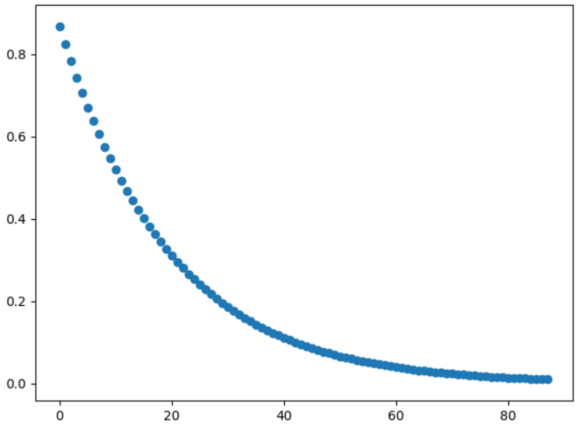
1.1 Question 1

The optimal policy π^* is easy to find because their is only 3 $(state, action)$ that have a reward. And their is only three steps.

$$\pi^* = [1, 1, 2]$$

1.2 Question 2

Figure 1: $\| v^k - v^* \|_\infty$



The value iteration find the same policy π^* and:

$$v^* = [15.204, 16.361, 17.819]$$

1.3 Question 3

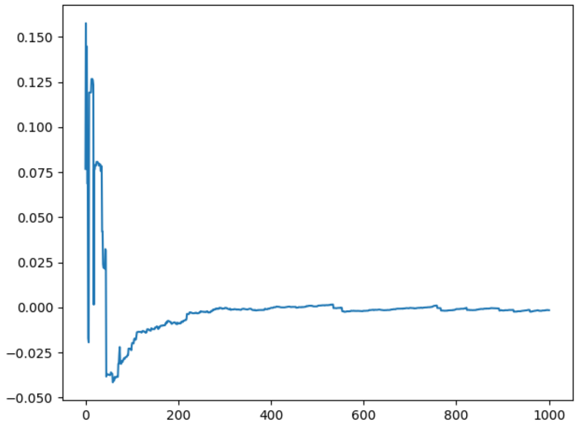
The exact policy iteration returned the same policy.
To compare both algorithm we used the *timeit* module of python.

	Mean of 100 runs
VI	0.00208620
PI	0.00179925

2 Reinforcement Learning

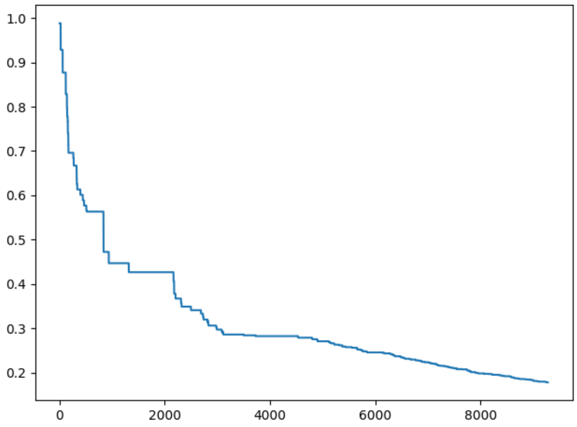
2.1 Question 4

Figure 2: $J_n - J^\pi$



2.2 Question 5

Figure 3: $\| v^k - v^* \|_\infty$


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
2000
4000
6000
8000