RL Assignment2

Aradhya PHD17008

September 2019

1

Please find in repo. Assuming that we receive unit reward for collection at a given time step.

2

Please find in repo.

Q2. Ax = b. where A (25x25) from each state going to every other state in a grid of size 5 is the action space and x is the policy. b is the reward obtained for the policy.

3

Q3. Please find in repo.
Q4. As explained in Q2. We select the A matrix to be defined for all actions i.e 25*4 = 100. The vector b is of size 100x1.

Q7. In Q7 we observe graphs similar to the ones shown in the book. We see the positive movement from one location and negative from another. The value observed is similar to the value function shown in the book.

4 Outputs

5

```
[[ 3.3  8.8  4.4  5.3  1.5]
 [ 1.5  3.   2.3  1.9  0.5]
 [ 0.1  0.7  0.7  0.4 -0.4]
 [-1.  -0.4 -0.4 -0.6 -1.2]
 [-1.9 -1.3 -1.2 -1.4 -2. ]]
```

Figure 1: Q2. Value function

```
Vpi
[[21.97748529 24.4194281  20.97748529 18.4501845  15.60516605]
 [19.77973676 21.97748529 18.87973676 16.60516605 14.04464945]
 [17.80176308 19.77973676 16.99176308 14.94464945 12.6401845 ]
 [16.02158677 17.80176308 15.29258677 13.4501845  11.37616605]
 [14.4194281  16.02158677 13.7633281  12.10516605 10.23854945]]
```

Figure 2: Q4. Obtained value function

```
[1] | [2] | [0] | [0] | [0] |
[1, 2] | [2] | [0] | [0] | [0] |
[1, 2] | [2] | [0] | [0] | [0] |
[1, 2] | [2] | [0] | [0] | [0] |
[1, 2] | [2] | [0] | [0] | [0] |
```

Figure 3: Q4. A sample Policy obtained

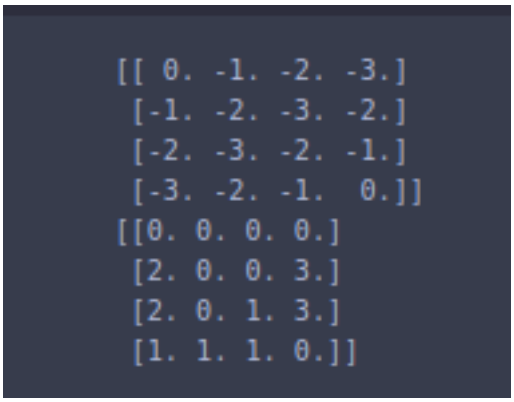

Figure 4: Q6. Value function and sample Policy obtained by value iteration

and sample Policy obtained by Policy iteration

Figure 5: Caption

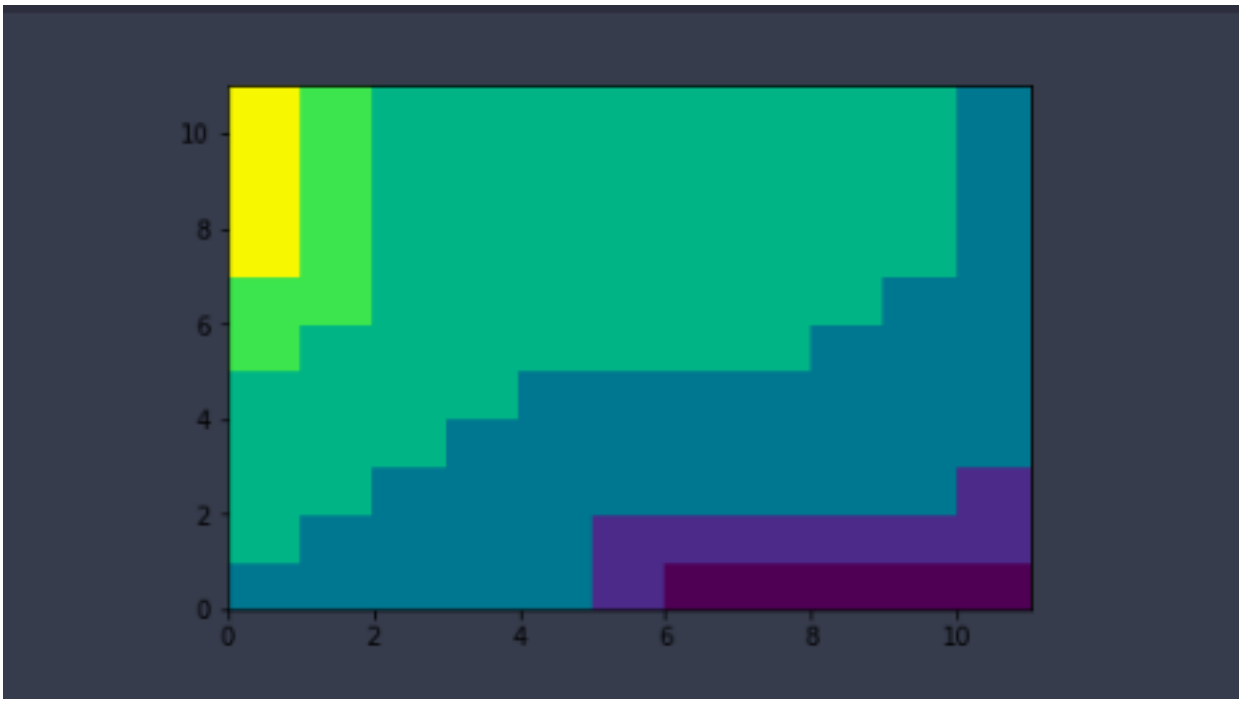

Figure 6: Q7. Policy obtained

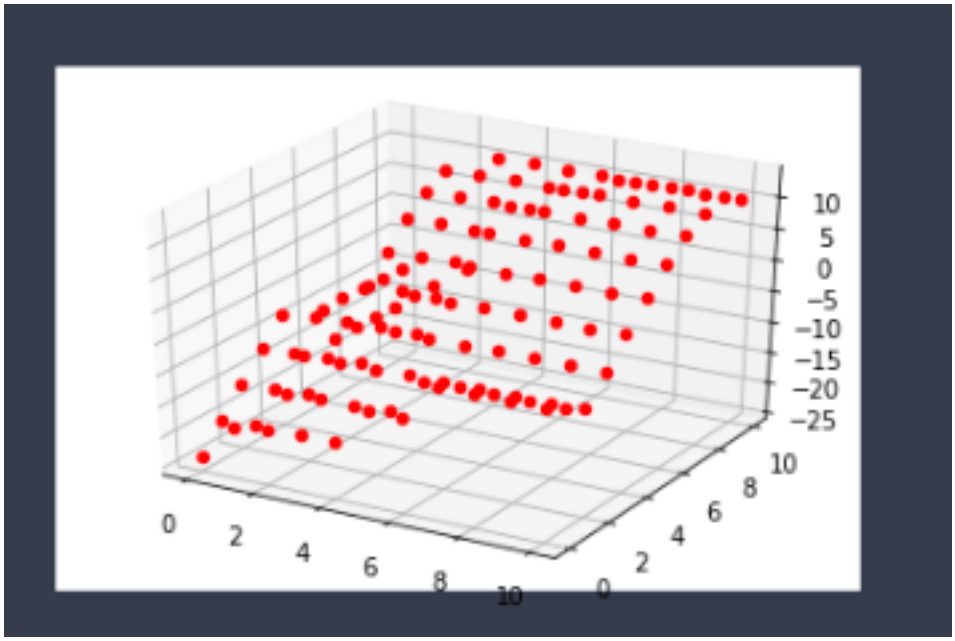

Figure 7: Q7. Value function obtained