

Assignment 1 REINFORCEMENT LEARNING

QUESTION 1:

Q1.pdf

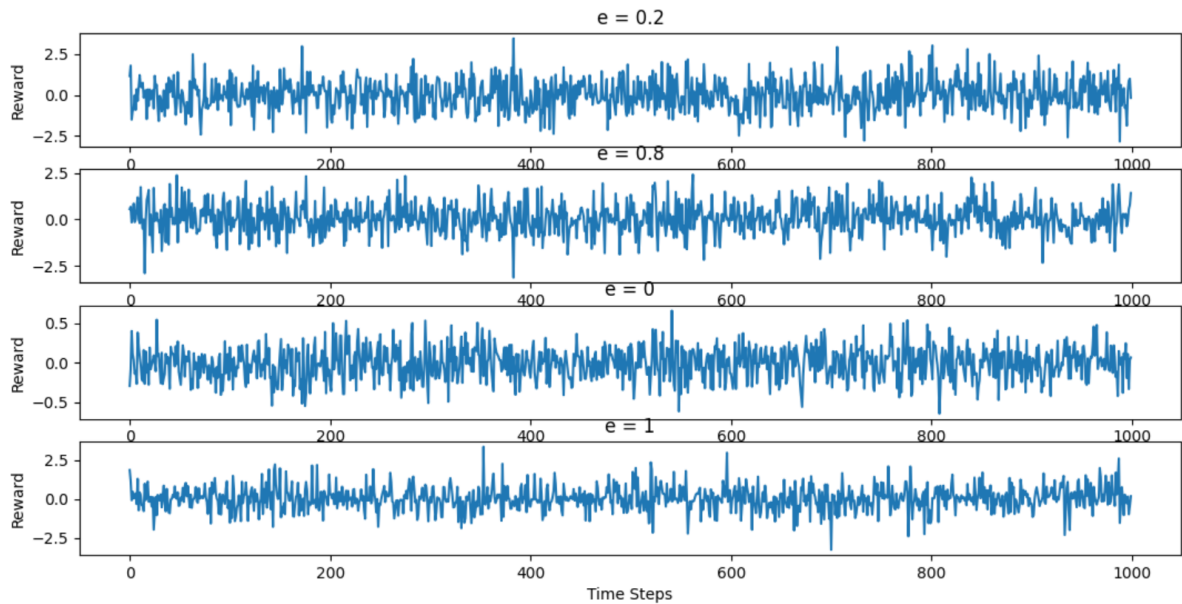
QUESTION 2:

Next page

QUESTION 2:

Q2.py

Figure 1



Visual Studio Code interface showing the Q2.py file and its execution output.

```
Q2.py
45 # (ii)
46 arr2 = Q2(0.8)
47 print(np.average(arr2))
48 plt.subplot(4, 1, 2)
49 plt.xlabel('Time Steps')
50 plt.ylabel('Reward')
51 plt.title('e = 0.8')
52 plt.plot(arr2)
53
54 # (iii)
55 arr3 = Q2(0)
56 print(np.average(arr3))
57 plt.subplot(4, 1, 3)
58 plt.xlabel('Time Steps')
```

Terminal Output:

```
Windows PowerShell
Copyright (C) Microsoft Corporation

Install the latest PowerShell for
...

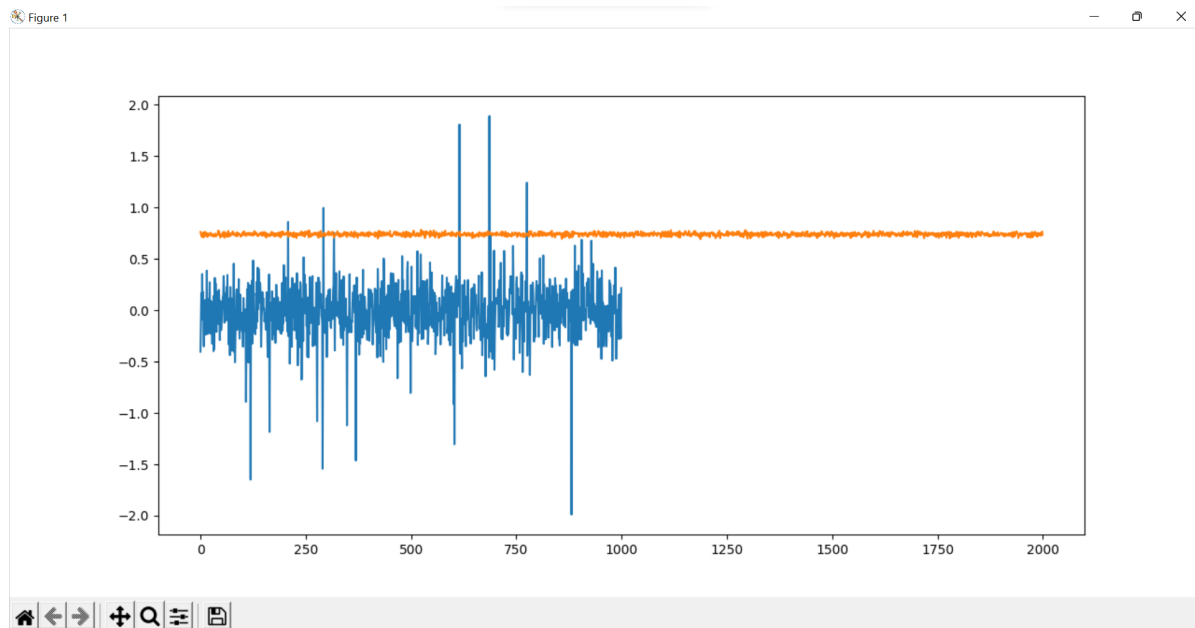
PS C:\Users\sahas\Desktop\MAB Proj>
ct/Q2.py"
0.10367071935049384
0.04473577794755956
-0.004234544087814507
0.05260691582571044
```

Avg Values

Figure 1 window showing the same four vertically stacked line plots as above, with the same axes and data series.

QUESTION 5:

Q5.py



Exercise 2.8:

Note: I have not got the exact graph as Fig. 2.4, but I can infer the following.

UCB Spikes

We can see a spike at the 11th step as the question is of 10 Arms.

So, after each arm is once iterated over, or each arm is once initialized, we now have the option to choose the best (most optimal) arm at the 11th episode.

Now at subsequent pulls, we see a fall because everything is kind of a reset. By this, I mean that we now cannot choose the optimal reward and have to start exploring again like we did at the start.

Also, if c is more the spike will be more as we start to explore.