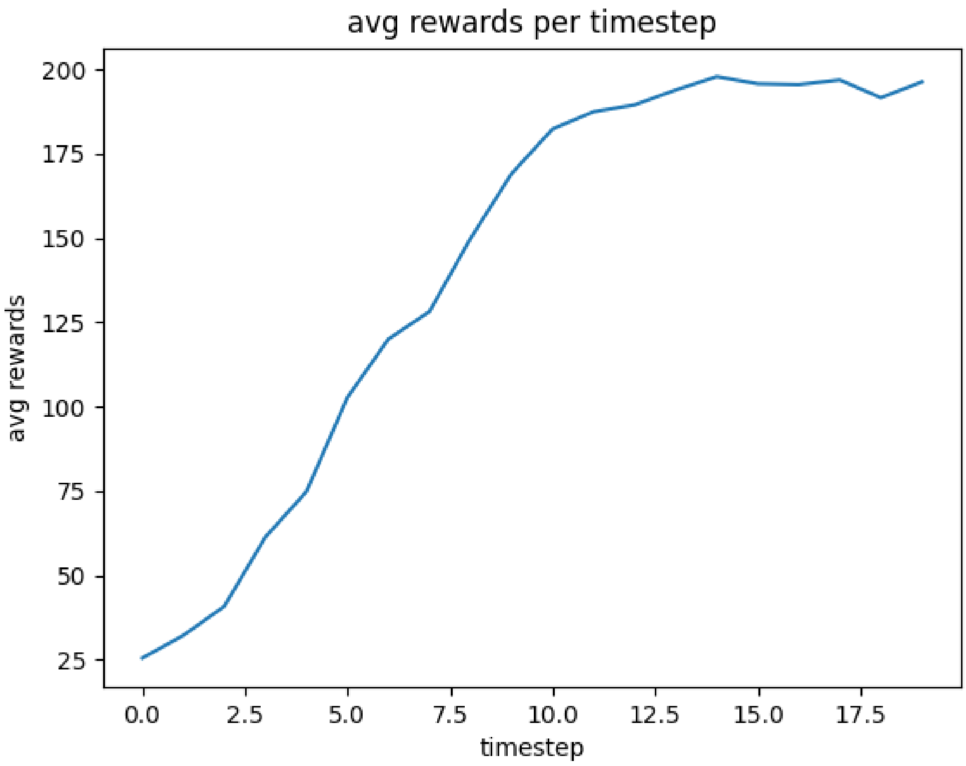
CS 4789 PA 4 Discussion

Chad Yu

May 2, 2024

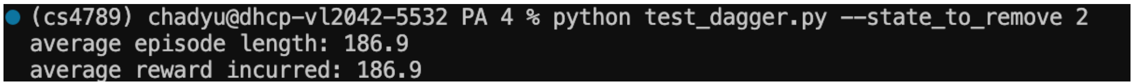
1 Question 1

The plot for the average rewards for each time step is shown below. We can see that indeed the average reward at T = 15 is around 190, as expected.

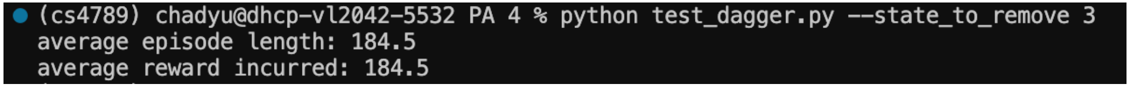


2 Question 2

First, as a baseline, we observed an average reward of 200.0 when the learned policy did not have any states removed, which achieves a reward of at least 190. In the testing with removed states, we observed that states 0 and 1 were able to be removed without compromising any performance, while removing states 2 or 3 reduced the performance of the agent approximately the same. While testing learned policies with states 0 or 1 removed, we never saw the reward dip below 200.0, but if we removed states 2 or 3, some tests on learned policies would go below our baseline of 200.0, and even below the expected baseline of 190, as shown below.



```
● (cs4789) chadyu@dhcp-vl2042-5532 PA 4 % python test_dagger.py --state_to_remove 2
  average episode length: 186.9
  average reward incurred: 186.9
```



```
● (cs4789) chadyu@dhcp-vl2042-5532 PA 4 % python test_dagger.py --state_to_remove 3
  average episode length: 184.5
  average reward incurred: 184.5
```

Our visual results aligned with these numerical results too; when rendering the Cartpole environment, for states 0 or 1 removed, the Cartpole was able to stay still upright, with none to very little oscillation, while with states 2 or 3 removed, there were clear swaying and swooping movements of the Cartpole. Even though there were some observations of reward 200.0 across different learned policies or tests with states 2 or 3 removed, these unstable movements were still present, further confirming our numerical results, and the fact that removing states 0 or 1 does not compromise performance, while removing states 2 or 3 evidently does so.