DeepQ

June 4, 2018

```
In [1]: %run custom_cartpole.py

WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
update steps =  1000
-------------------------------------
| % time spent exploring | 59       |
| episodes               | 200      |
| mean episode reward    | 21.6     |
| steps                  | 4177     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 400      |
| mean episode reward    | 155      |
| steps                  | 24822    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 600      |
| mean episode reward    | 176      |
| steps                  | 62251    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 800      |
| mean episode reward    | 182      |
| steps                  | 96505    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 1000     |
| mean episode reward    | 200      |
| steps                  | 136505   |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
update steps =  250
-------------------------------------
```

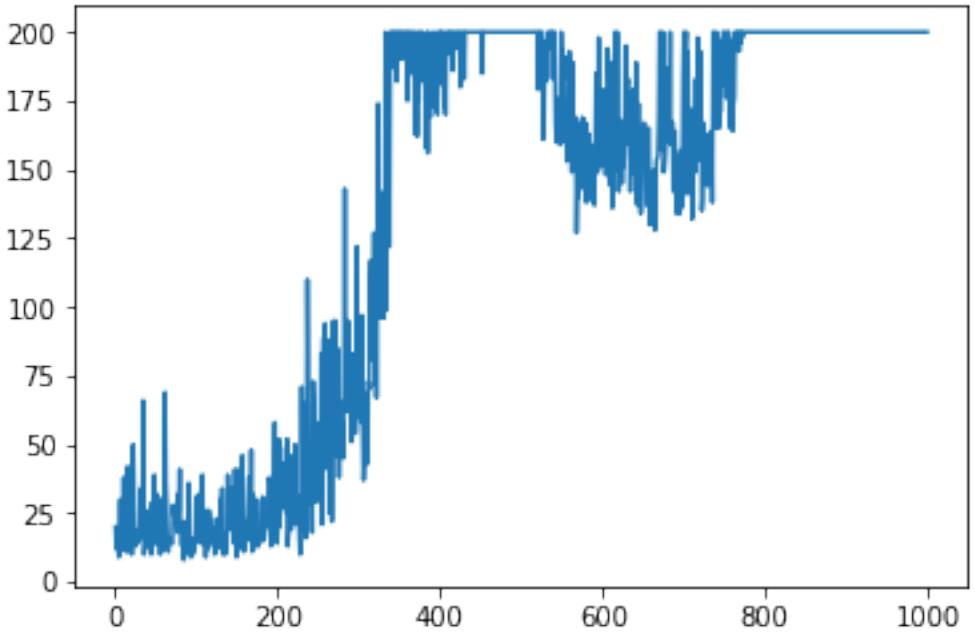
```
| % time spent exploring | 34       |
| episodes               | 200      |
| mean episode reward    | 39.2     |
| steps                  | 6666     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 400      |
| mean episode reward    | 173      |
| steps                  | 35016    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 600      |
| mean episode reward    | 70       |
| steps                  | 51855    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 800      |
| mean episode reward    | 19.9     |
| steps                  | 55734    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 1000     |
| mean episode reward    | 193      |
| steps                  | 91816    |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
update steps =  50
-------------------------------------
| % time spent exploring | 61       |
| episodes               | 200      |
| mean episode reward    | 18       |
| steps                  | 3917     |
-------------------------------------
-------------------------------------
| % time spent exploring | 35       |
| episodes               | 400      |
| mean episode reward    | 12.3     |
| steps                  | 6632     |
-------------------------------------
-------------------------------------
| % time spent exploring | 13       |
| episodes               | 600      |
| mean episode reward    | 10.5     |
| steps                  | 8803     |
```

```
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 800      |
| mean episode reward    | 9.4      |
| steps                  | 10746    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 1000     |
| mean episode reward    | 9.5      |
| steps                  | 12635    |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
update steps =  1
-------------------------------------
| % time spent exploring | 60       |
| episodes               | 200      |
| mean episode reward    | 18.1     |
| steps                  | 3981     |
-------------------------------------
-------------------------------------
| % time spent exploring | 34       |
| episodes               | 400      |
| mean episode reward    | 12.4     |
| steps                  | 6715     |
-------------------------------------
-------------------------------------
| % time spent exploring | 12       |
| episodes               | 600      |
| mean episode reward    | 10.3     |
| steps                  | 8902     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 800      |
| mean episode reward    | 9.4      |
| steps                  | 10820    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 1000     |
| mean episode reward    | 9.4      |
| steps                  | 12709    |
-------------------------------------
```

```
In [2]: %matplotlib inline
```

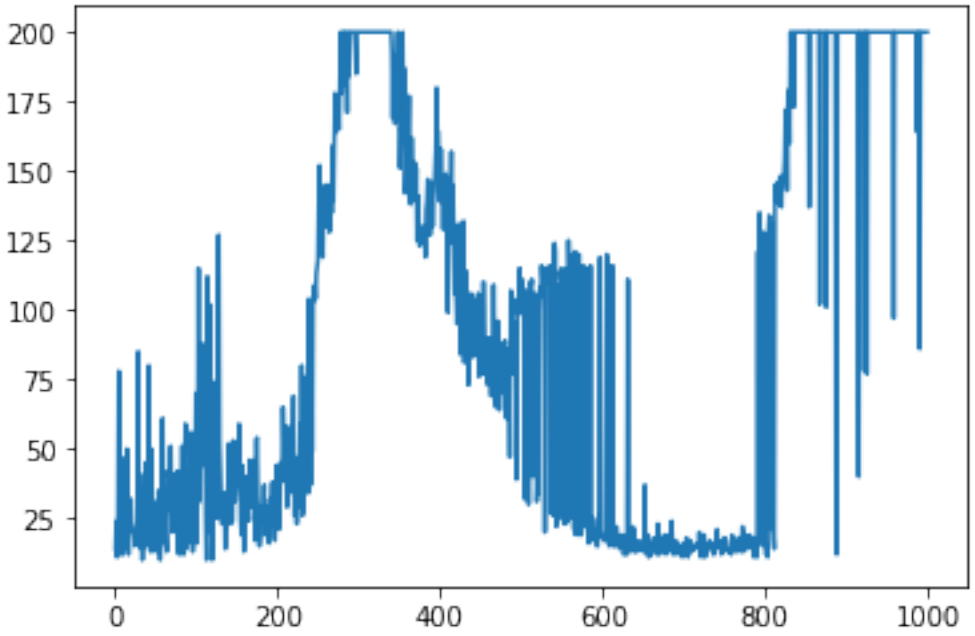
```
import matplotlib.pyplot as plt
x = np.linspace(1,999,999)
plt.plot(x,reward_history[0,0:999])
```

Out[2]: [<matplotlib.lines.Line2D at 0x21617bf0828>]



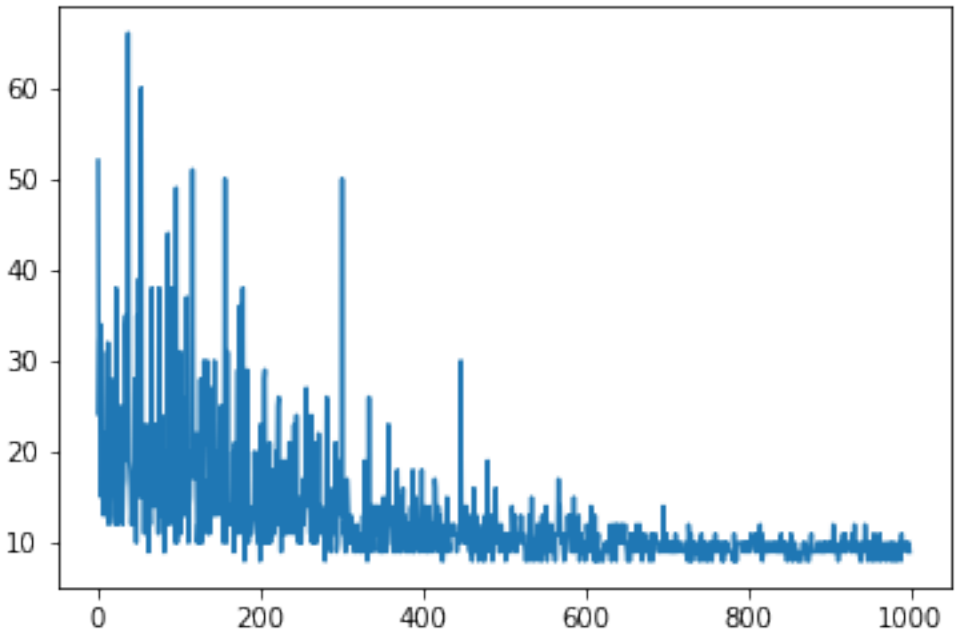
In [3]: `plt.plot(x,reward_history[1,0:999])`

Out[3]: [<matplotlib.lines.Line2D at 0x21617c8c9e8>]



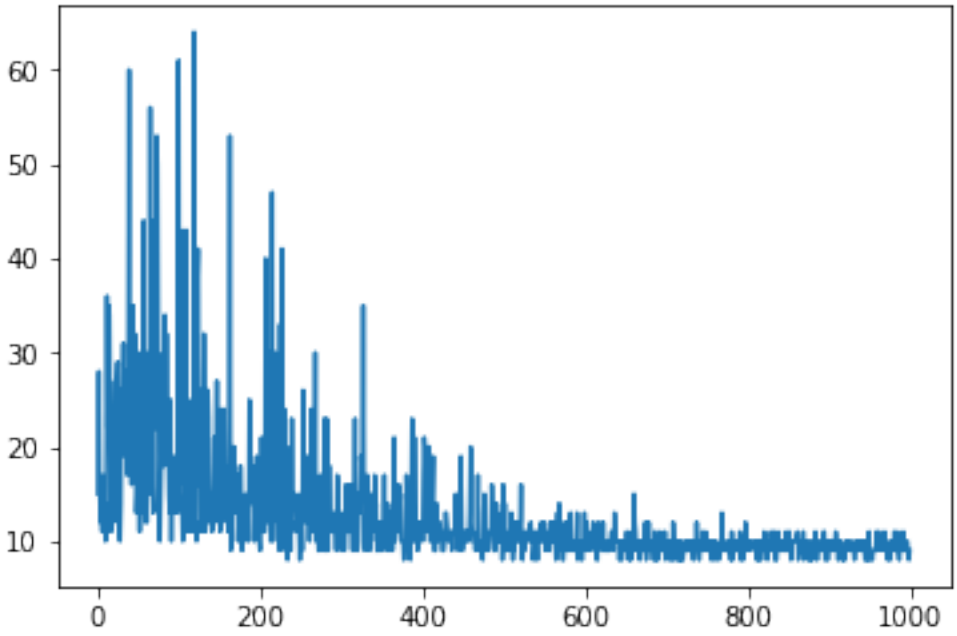

```
In [4]: plt.plot(x,reward_history[2,0:999])
```

```
Out[4]: [<matplotlib.lines.Line2D at 0x21617f26748>]
```



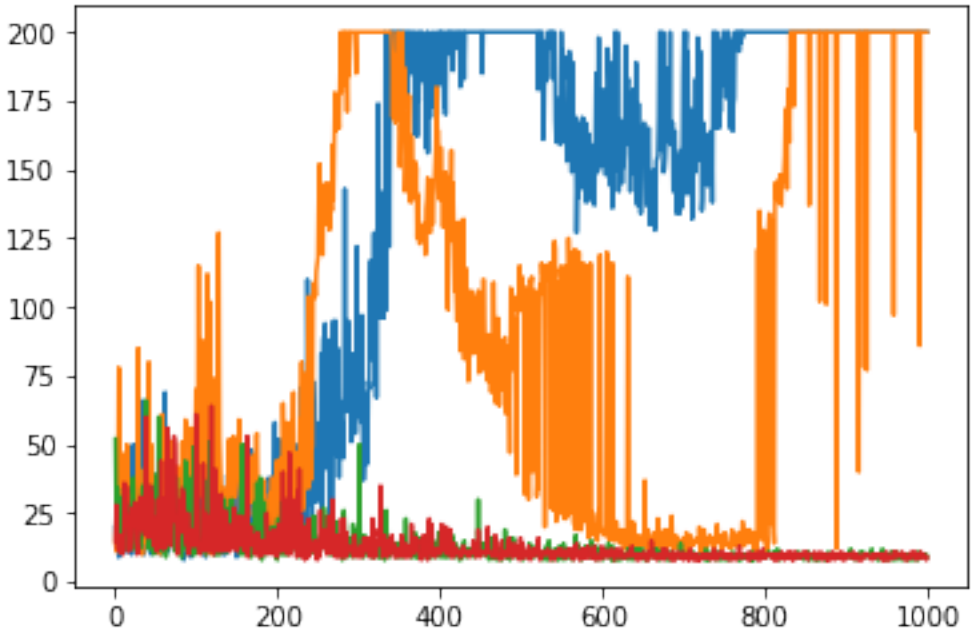
```
In [5]: plt.plot(x,reward_history[3,0:999])

Out[5]: [<matplotlib.lines.Line2D at 0x21617f88588>]
```




```
In [6]: plt.plot(x,reward_history[0,0:999],x,reward_history[1,0:999],x,reward_history[2,0:999],x

Out[6]: [<matplotlib.lines.Line2D at 0x21617fe2c88>,
         <matplotlib.lines.Line2D at 0x21617fe2e48>,
         <matplotlib.lines.Line2D at 0x21617fec6a0>,
         <matplotlib.lines.Line2D at 0x21617fecb00>]
```



The target network is similar to $\overline{V}(s')$ *in the value iteration. We keep it constant while updating the weights of* $Q-$ *network. Reducing the update steps will potentially increase the speed of convergence. However, if the update becomes too fr*

```
In [1]: %run custom_cartpole2.py

WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
mini batch size =  32
-------------------------------------
| % time spent exploring | 59       |
| episodes               | 200      |
| mean episode reward    | 20       |
| steps                  | 4136     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 400      |
| mean episode reward    | 107      |
| steps                  | 18696    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 600      |
| mean episode reward    | 171      |
| steps                  | 52567    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
```

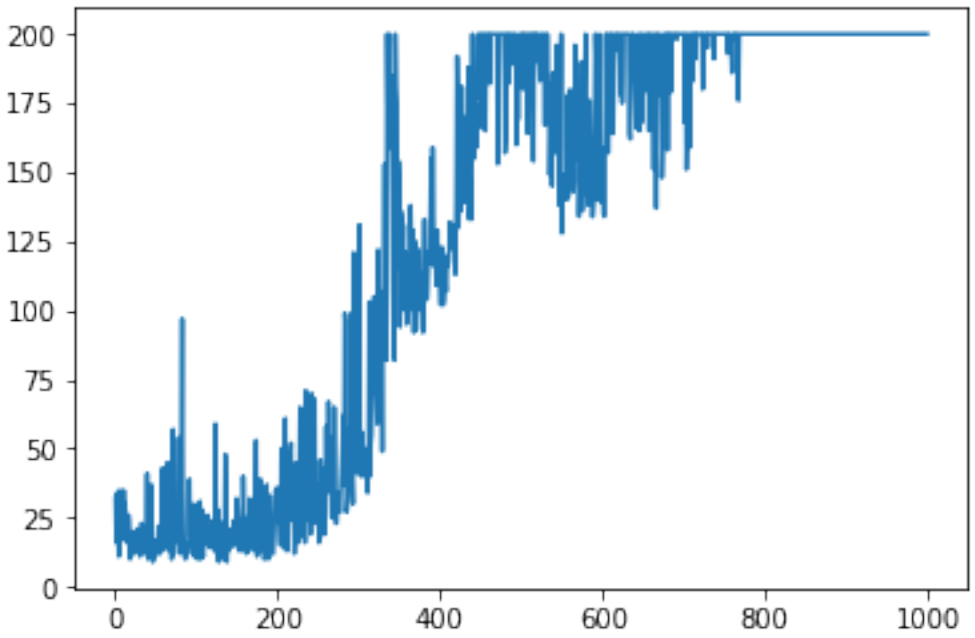
```
| episodes                | 800      |
| mean episode reward     | 197      |
| steps                   | 91113    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 1000     |
| mean episode reward     | 200      |
| steps                   | 131113   |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
mini batch size =  15
-------------------------------------
| % time spent exploring | 52       |
| episodes                | 200      |
| mean episode reward     | 26.2     |
| steps                   | 4807     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 400      |
| mean episode reward     | 179      |
| steps                   | 30209    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 600      |
| mean episode reward     | 187      |
| steps                   | 68166    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 800      |
| mean episode reward     | 188      |
| steps                   | 104415   |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 1000     |
| mean episode reward     | 200      |
| steps                   | 144226   |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
mini batch size =  5
-------------------------------------
| % time spent exploring | 60       |
| episodes                | 200      |
| mean episode reward     | 18.3     |
```

```
| steps                  | 4019     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 400      |
| mean episode reward    | 54.5     |
| steps                  | 12014    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 600      |
| mean episode reward    | 158      |
| steps                  | 34276    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 800      |
| mean episode reward    | 198      |
| steps                  | 73623    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 1000     |
| mean episode reward    | 138      |
| steps                  | 105870   |
-------------------------------------
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.float32'>. Please provide explicit dtyp
mini batch size =  1
-------------------------------------
| % time spent exploring | 59       |
| episodes               | 200      |
| mean episode reward    | 20.7     |
| steps                  | 4157     |
-------------------------------------
-------------------------------------
| % time spent exploring | 9        |
| episodes               | 400      |
| mean episode reward    | 28.5     |
| steps                  | 9261     |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes               | 600      |
| mean episode reward    | 35.4     |
| steps                  | 14658    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
```

```
| episodes                | 800      |
| mean episode reward     | 81.4     |
| steps                   | 29081    |
-------------------------------------
-------------------------------------
| % time spent exploring | 2        |
| episodes                | 1000     |
| mean episode reward     | 147      |
| steps                   | 53884    |
-------------------------------------
```

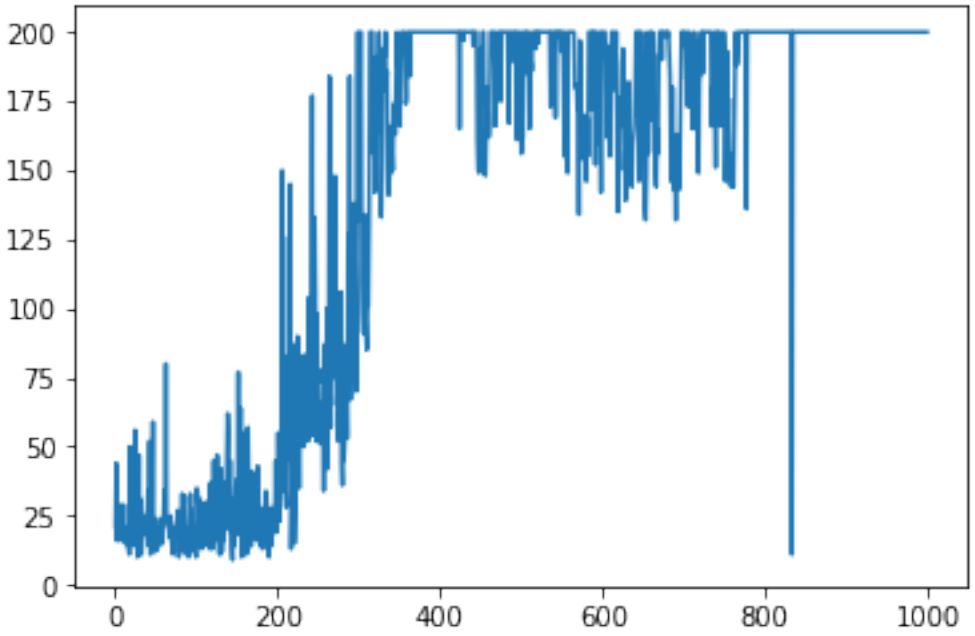
```
In [2]: %matplotlib inline
        import matplotlib.pyplot as plt
        x = np.linspace(1,999,999)
        plt.plot(x,reward_history[0,0:999])
```

```
Out[2]: [<matplotlib.lines.Line2D at 0x11d4e96a8d0>]
```



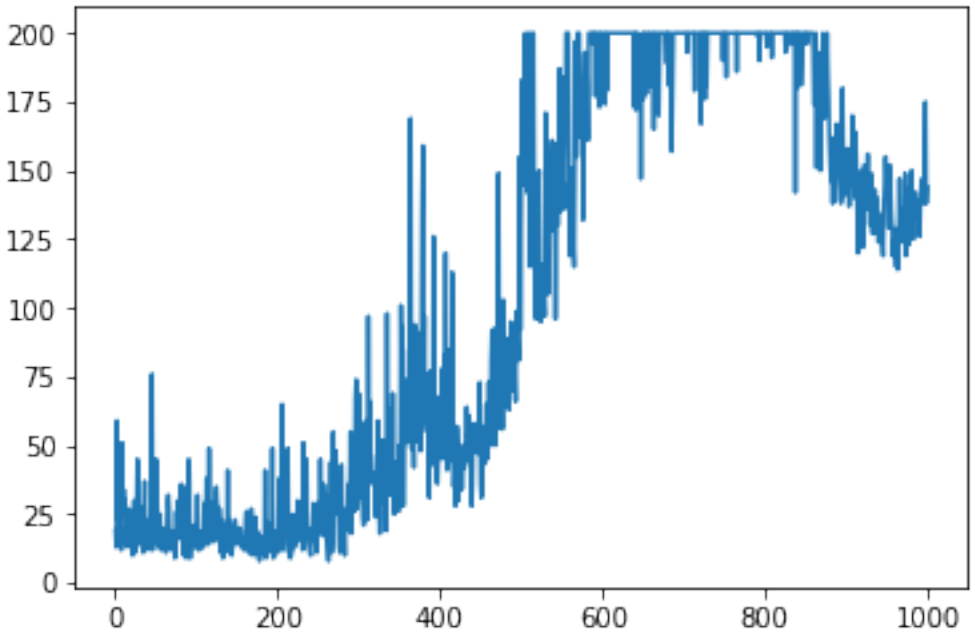

```
In [3]: plt.plot(x,reward_history[1,0:999])
```

```
Out[3]: [<matplotlib.lines.Line2D at 0x11d4eb4aa20>]
```



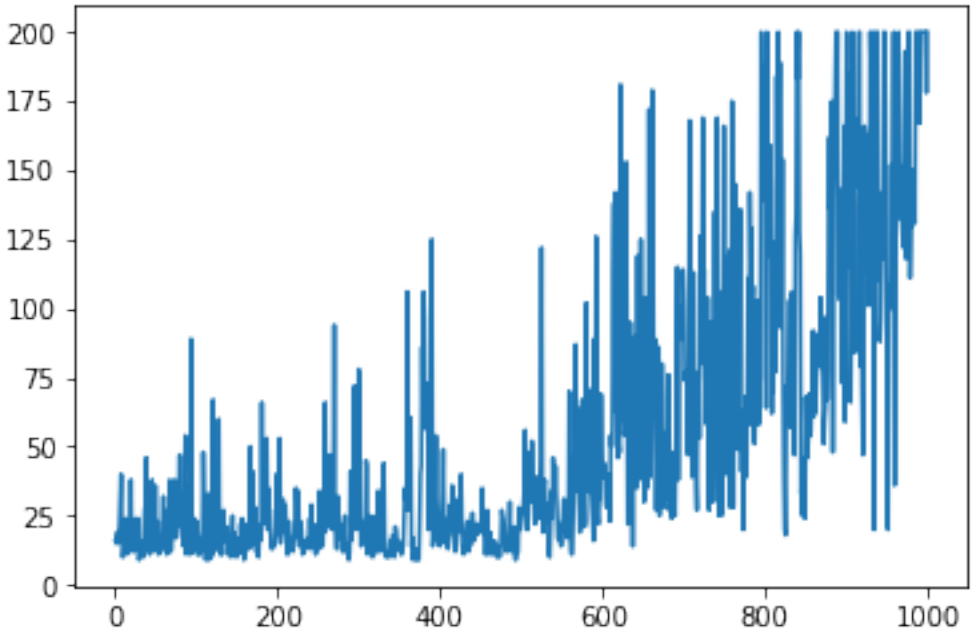
In [4]: `plt.plot(x,reward_history[2,0:999])`

Out[4]: `[<matplotlib.lines.Line2D at 0x11d4ebb5b38>]`



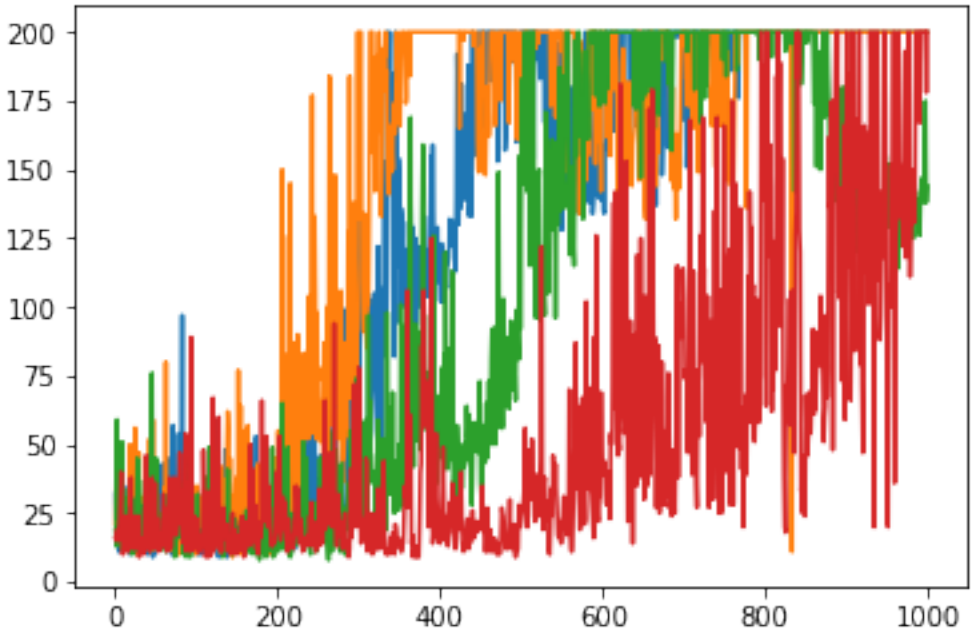
```
In [5]: plt.plot(x,reward_history[3,0:999])

Out[5]: [<matplotlib.lines.Line2D at 0x11d4ec1d7b8>]
```




```
In [6]: plt.plot(x,reward_history[0,0:999],x,reward_history[1,0:999],x,reward_history[2,0:999],x

Out[6]: [<matplotlib.lines.Line2D at 0x11d4fe61da0>,
         <matplotlib.lines.Line2D at 0x11d4fe61f60>,
         <matplotlib.lines.Line2D at 0x11d4fe6b7b8>,
         <matplotlib.lines.Line2D at 0x11d4fe6bc18>]
```



By comparing the results, we observe that when decreasing the size of the mini-batches, the variance increases. It is because when using a large batch, the gradient can be averaged among the samples, thus reducing the variance. In addition, the speed of converging is slower when decreasing the batch size. It is becuse in each episode, the larger the batch size is, the more experience the system can obtain from training history, and it should learn faster.