ACROBOT

1. SARSA:

Step 499 @ Episode 1/1000 (0.0)Episode finished after 500 timesteps

Step 499 @ Episode 2/1000 (-500.0)Episode finished after 500 timesteps

Step 499 @ Episode 3/1000 (-500.0)Episode finished after 500 timesteps

Step 499 @ Episode 4/1000 (-500.0)Episode finished after 500 timesteps

Step 499 @ Episode 5/1000 (-500.0)Episode finished after 500 timesteps

Step 498 @ Episode 6/1000 (-500.0)Episode finished after 499 timesteps

….

Step 107 @ Episode 995/1000 (-93.0)Episode finished after 108 timesteps

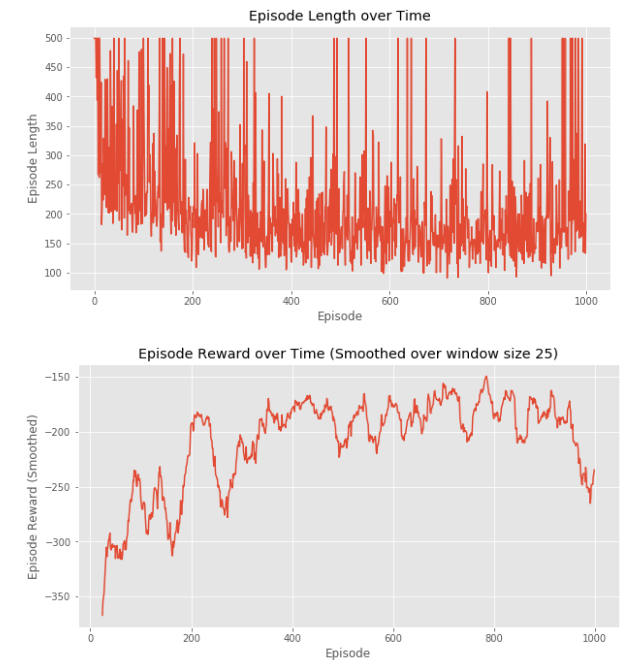
Step 499 @ Episode 996/1000 (-107.0)Episode finished after 500 timesteps

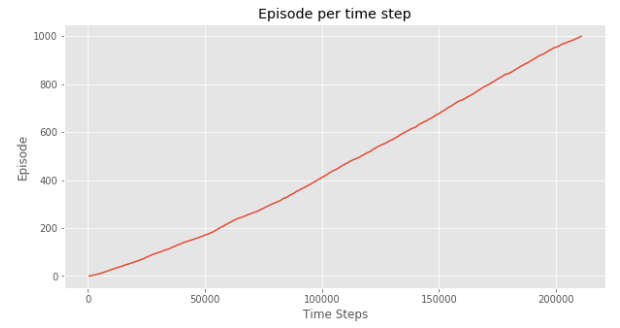
Step 137 @ Episode 997/1000 (-500.0)Episode finished after 138 timesteps

Step 121 @ Episode 998/1000 (-137.0)Episode finished after 122 timesteps

Step 103 @ Episode 999/1000 (-121.0)Episode finished after 104 timesteps

Step 142 @ Episode 1000/1000 (-103.0)Episode finished after 143 timesteps





1. Q learning with Value Function Approximation :

Step 499 @ Episode 1/1000 (0.0)Episode finished after 500 timesteps

Step 499 @ Episode 2/1000 (-500.0)Episode finished after 500 timesteps

Step 499 @ Episode 3/1000 (-500.0)Episode finished after 500 timesteps

Step 499 @ Episode 4/1000 (-500.0)Episode finished after 500 timesteps

Step 432 @ Episode 5/1000 (-500.0)Episode finished after 433 timesteps

Step 499 @ Episode 6/1000 (-432.0)Episode finished after 500 timesteps

…..

Step 135 @ Episode 995/1000 (-185.0)Episode finished after 136 timesteps

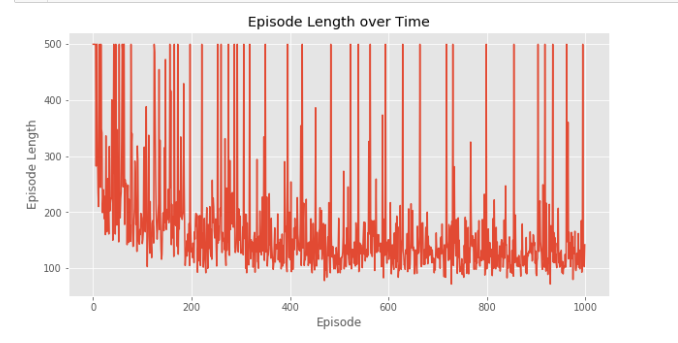
Step 163 @ Episode 996/1000 (-135.0)Episode finished after 164 timesteps

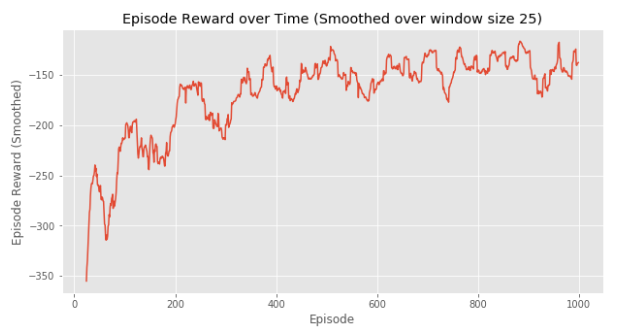
Step 189 @ Episode 997/1000 (-163.0)Episode finished after 190 timesteps

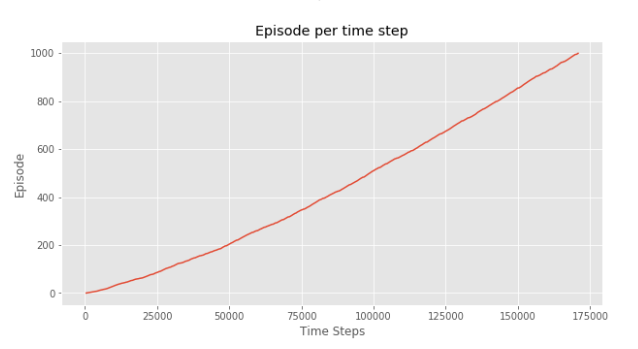
Step 319 @ Episode 998/1000 (-189.0)Episode finished after 320 timesteps

Step 133 @ Episode 999/1000 (-319.0)Episode finished after 134 timesteps

Step 200 @ Episode 1000/1000 (-133.0)Episode finished after 201 timesteps







3)DQN:  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

flatten\_1 (Flatten) (None, 6) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 16) 112

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

activation\_1 (Activation) (None, 16) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 16) 272

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

activation\_2 (Activation) (None, 16) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_3 (Dense) (None, 16) 272

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

activation\_3 (Activation) (None, 16) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_4 (Dense) (None, 3) 51

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

activation\_4 (Activation) (None, 3) 0

=================================================================

Total params: 707

Trainable params: 707

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

None

Training for 10000 steps ...

500/10000: episode: 1, duration: 10.493s, episode steps: 500, steps per second: 48, episode reward: -500.000, mean reward: -1.000 [-1.000, -1.000], mean action: 1.062 [0.000, 2.000], mean observation: 0.309 [-3.540, 3.879], loss: 0.043475, mean\_absolute\_error: 1.384563, mean\_q: -1.893409

1000/10000: episode: 2, duration: 8.341s, episode steps: 500, steps per second: 60, episode reward: -500.000, mean reward: -1.000 [-1.000, -1.000], mean action: 0.980 [0.000, 2.000], mean observation: 0.170 [-7.574, 7.768], loss: 0.025416, mean\_absolute\_error: 3.902266, mean\_q: -5.740516

1500/10000: episode: 3, duration: 8.341s, episode steps: 500, steps per second: 60, episode reward: -500.000, mean reward: -1.000 [-1.000, -1.000], mean action: 1.024 [0.000, 2.000], mean observation: 0.219 [-7.272, 7.517], loss: 0.071810, mean\_absolute\_error: 6.456538, mean\_q: -9.486332

….

9793/10000: episode: 32, duration: 5.844s, episode steps: 351, steps per second: 60, episode reward: -350.000, mean reward: -0.997 [-1.000, 0.000], mean action: 1.085 [0.000, 2.000], mean observation: 0.497 [-9.592, 11.917], loss: 1.110230, mean\_absolute\_error: 28.333006, mean\_q: -41.540363

9962/10000: episode: 33, duration: 2.830s, episode steps: 169, steps per second: 60, episode reward: -168.000, mean reward: -0.994 [-1.000, 0.000], mean action: 0.935 [0.000, 2.000], mean observation: 0.228 [-6.382, 6.697], loss: 1.097991, mean\_absolute\_error: 28.561075, mean\_q: -41.866108

done, took 168.993 seconds

Testing for 5 episodes ...

Testing for 5 episodes ...

Episode 1: reward: -86.000, steps: 87

Episode 2: reward: -93.000, steps: 94

Episode 3: reward: -92.000, steps: 93

Episode 4: reward: -75.000, steps: 76

Episode 5: reward: -77.000, steps: 78

**Also tried 50000, the results are similar….**