09/11

Code working for Acrobot with height as reward

Result: Not balancing the acrobot

09/13

1. Experiment with the Pendulum environment

Result: Successfully balance the pendulum

2. Modify the acrobat reward function to include the cost of thetadot

Result: Not balancing, swings over constantly

Concern: dt too large

09/20

Add functionality to change available torque and dt based on a factor

Result: Not balancing, swings over, dt changes doesn’t seem to have much effect

Concern: Not enough torque

09/25

Change dt without change torque and start near straight up

Result: no distinctive difference between the results; does well in some initial conditions but fails more often

Concern: not enough torque options, not large enough neural network, not enough torque, not enough training, overfitting

09/27

FIX dt TO 2 FOR FOLLOWING EXPERIMENTS

1. Change neural network to [128, 128]

Result: significant improvement of score and Q value

2. Change stacking of history to one state

Result: significant improvement of score and Q value

3. Larger torque (by 2, by 4)

Result: Not strictly tested

4. More options with larger torque

Result: Not helping if not hurting

Concern: action space size might not be properly changed, MIGHT NEED TO RERUN

5. Rerun 09/25 experiments with longer training (change stopping condition)

DID NOT RERUN BY OCT 1ST

6. Add dropout on the second layer of [32, 32] network

Result: No significant improvement

10/01

Run no stacking of history with large neural networks: [128,128],[256,256],[512,512]

Result: no significant improvement as neural networks get larger

10/08

Run no stacking of history with deeper neural network

Result: deeper neural network does have more stable performance but not significant improvement

Run pendulum with action taken out of reward function

Result: No significant difference but it diverges faster

10/14

Use frames lasted as the reward

Result: slightly more stable

10/30 (256,256) neural network

Wider range of actions

Result: Very unstable

Starting straight up without randomness

Result: steady increase and then fluctuates around 150

11/03

Slower decay of epsilon

Result: