Continuous Control Project report

The environment

In this environment, a double-jointed arm can move to target locations. A reward of +0.1 is provided for each step that the agent's hand is in the goal location. Thus, the goal of your agent is to maintain its position at the target location for as many time steps as possible.

The observation space consists of 33 variables corresponding to position, rotation, velocity, and angular velocities of the arm. Each action is a vector with four numbers, corresponding to torque applicable to two joints. Every entry in the action vector should be a number between -1 and 1.

The task is episodic, and in order to solve the environment, the agent must get an average score of +30 over 100 consecutive episodes.

Solution

I have used actor-critic method for solving single agent version of the problem. I have used the DDPG algorithm to learn a neural network based function approximators for actor and critic. The paper describing this algorithm can be found here.

Hyperparameters used are:

BUFFER_SIZE = int(1e6) # replay buffer size

BATCH_SIZE = 512   # minibatch size

GAMMA = 0.99   # discount factor

TAU = 1e-3   # for soft update of target parameters

LR_ACTOR = 3e-4   # learning rate of the actor

LR_CRITIC = 3e-3   # learning rate of the critic

WEIGHT_DECAY = 0   # L2 weight decay

UPDATE_EVERY = 4   # how often to update the network

Most of the code is reused from this repository: https://github.com/udacity/deep-reinforcement-learning/tree/master/ddpg-pendulum .

The below plot shows how agents average score changes with each episode:
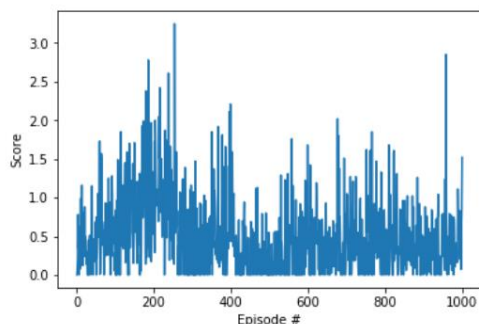
The environment was not solved even in 1000 episodes. Below is the average score after every 100 episodes:

```
Episode 100     Average Score: 0.46
Episode 200     Average Score: 0.78
Episode 300     Average Score: 0.84
Episode 400     Average Score: 0.52
Episode 500     Average Score: 0.30
Episode 600     Average Score: 0.33
Episode 700     Average Score: 0.41
Episode 800     Average Score: 0.47
Episode 900     Average Score: 0.42
Episode 1000    Average Score: 0.41
```