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.

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:
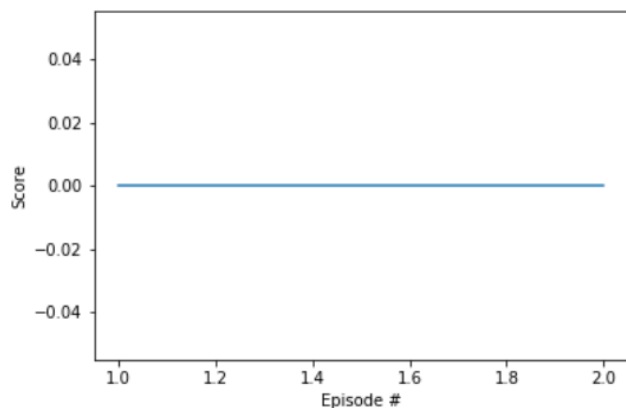
The environment was not solved even after more than 400 episodes. Below is the average score after every 100 episodes:

```
Episode 100     Average Score: 0.60
Episode 200     Average Score: 0.69
Episode 300     Average Score: 0.52
Episode 400     Average Score: 0.49
Episode 465     Average Score: 0.51
```