Continuous Control Project Report

Introduction

The reader is assumed to have basic knowledge on Reinforcement Learning and Deep Learning algorithms in particular DDPG.

This project aims at training an agent to control an arm to reach target locations. The task is episodic. The state space is a 33-dimensional space and the action space is a continuous 4-dimensional space where each number could be anything between -1 and 1. To achieve the task the agent must achieve an average score of +30 (averaged over all arms) over 100 episodes.

The specificity of the **DDPG algorithm**[1] will not be described in this report, the link to the related paper is given at the end of the page.

Future Work

As discussed below, I had some troubles implementing the first version of the environment, that's why my priority is to achieve this environment and then to implement other algorithms such as PPO or A3C to compare with DDPG and get further knowledge on this domain.

Method

My first tries were with the first environment where I didn't manage to get anything relevant. I spent hours and days changing hyperparameters in vain.Training when it was, was unstable and in 2000 episodes I didn't get anything useful. So that's why I decided to try my luck with the second one, in the hope that errors were not mine. I readapted the code I used for the first environment so that it takes into account the 20 arms, each arm filling the replay buffer with its own experience. Only one actor and one critic model was trained and used for those 20 arms. After correcting some bugs and tweaking hyperparameters I managed to achieve the task.

[1] https://arxiv.org/pdf/1509.02971.pdf

Implementation Characteristics

	DDPG
Memory Size	1 000 000
Batch Size	128
Discount Factor	0.99
Rate of Transfer for Soft Update	5e-4
Frequency of Update of The Target Network	Every time steps
Max steps in one episode	1000
Ornstein-Uhlenbeck Process Parameters	Mu=0 Theta=0.15 Sigma=0.15
Actor Architecture	Linear(33,128) ReLU Linear(128,128) ReLU Linear(128,128) ReLU Linear(128,4)
Critic Architecture	Linear(33,128) ReLU (Here output and action vector are concatenated) Linear(128+4,128) ReLU Linear(128,64) ReLU Linear(64,1)
Learning Rate Actor	5e-4
Learning Rate Critic	5e-4

Results

DDPG Agent average score over 100 episodes (average of the score of the 20 arms)

```
Episode 100     Average Score: 10.32
Episode 172     Average Score: 30.08
```