Project Report - Continuous Control

Introduction

Project Goal

The goal of this project is to solve the Unity ML-Agents Reacher 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.

Project Overview

The observation space consists of 33 variables corresponding to the 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 environment is considered solved if a reward of +30 is obtained for 100 consecutive episodes.

Algorithms

The algorithm used here is a Deep Deterministic Policy Gradient (DDPG). A DDPG is composed of two networks, actor and critic. During a step, the actor is used to estimate the best action, and the critic then uses this value as in a DDQN to evaluate the optimal action-value function.

Model architecture

The DDPG is composed of two networks, actor and critic. The details of these networks have been given below.

Actor Network

Layer	**Neurons**	**Activation**
Fully Connected Layer	(State Size, 256)	ReLU
Fully Connected Layer	(256, 128)	ReLU
Output Layer	(128, Action Size)	TanH

Critic Network

Layer	**Neurons**	**Activation**
Fully Connected Layer	(State Size, 256)	ReLU
Fully Connected Layer	(256 + Action Size, 128)	ReLU
Output Layer	(128, 1)	Linear

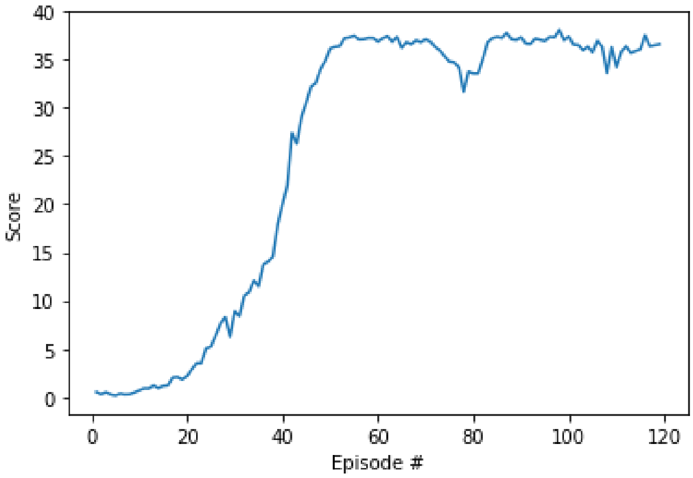
Hyper-parameters

- Learning Rate: 1e-4 (in both)
- Batch Size: 128
- Replay Buffer: 1e5

- Gamma: 0.99
- Tau: 1e-3
- Ornstein-Uhlenbeck noise parameters (0.15 theta and 0.2 sigma.)

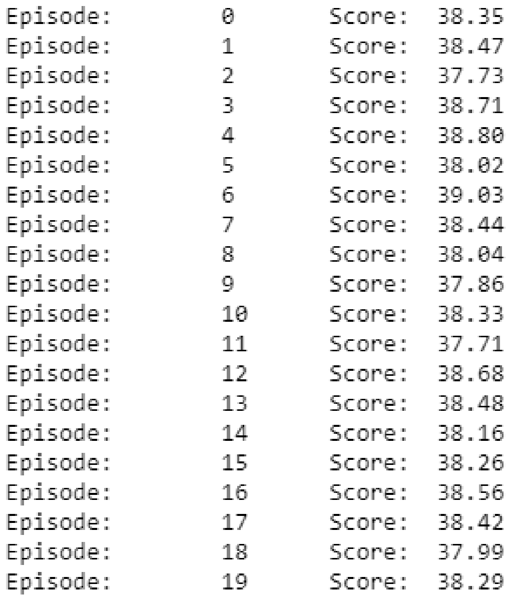
Results

Training



Training Log for the DDPG algorithm

Inference



```
Episode:        0       Score:  38.35
Episode:        1       Score:  38.47
Episode:        2       Score:  37.73
Episode:        3       Score:  38.71
Episode:        4       Score:  38.80
Episode:        5       Score:  38.02
Episode:        6       Score:  39.03
Episode:        7       Score:  38.44
Episode:        8       Score:  38.04
Episode:        9       Score:  37.86
Episode:        10      Score:  38.33
Episode:        11      Score:  37.71
Episode:        12      Score:  38.68
Episode:        13      Score:  38.48
Episode:        14      Score:  38.16
Episode:        15      Score:  38.26
Episode:        16      Score:  38.56
Episode:        17      Score:  38.42
Episode:        18      Score:  37.99
Episode:        19      Score:  38.29
```

Inference score of 20 episodes for the DDPG algorithm

Future Work

I plan to implement the D4PG and test both the algorithms (i.e., D4PG and DDPG) in both environments, i.e., reacher and the crawler. The goal is to find when and where each of the algorithms has the best performance.