

DRLND Project 2: Continuous Control

michaelfitzke

August 2019

1 Overview

This project utilised the architecture for solving the DDPG-Bipedal Udacity project repo. coding exercise as outlined in the solution provided by udacity.

2 State and Action Space

The State Space consists of 33 Variables. Each action is a vector of four numbers.

3 Model

We use the DDPG algorithm following [1].

3.1 Hyperparameters

The environment was solved with the following hyperparameters

- $BUFFER_{SIZE} = int(1e6)$
- $BATCH_{SIZE} = 128$
- $\gamma = 0.99$
- $\tau = 1e - 3$
- $LR_{ACTOR} = 1e - 4$
- $LR_{CRITIC} = 1e - 4$
- $WEIGHT_{DECAY} = 0.0$

4 Result

An average score over 30 was reached in 195 episodes.

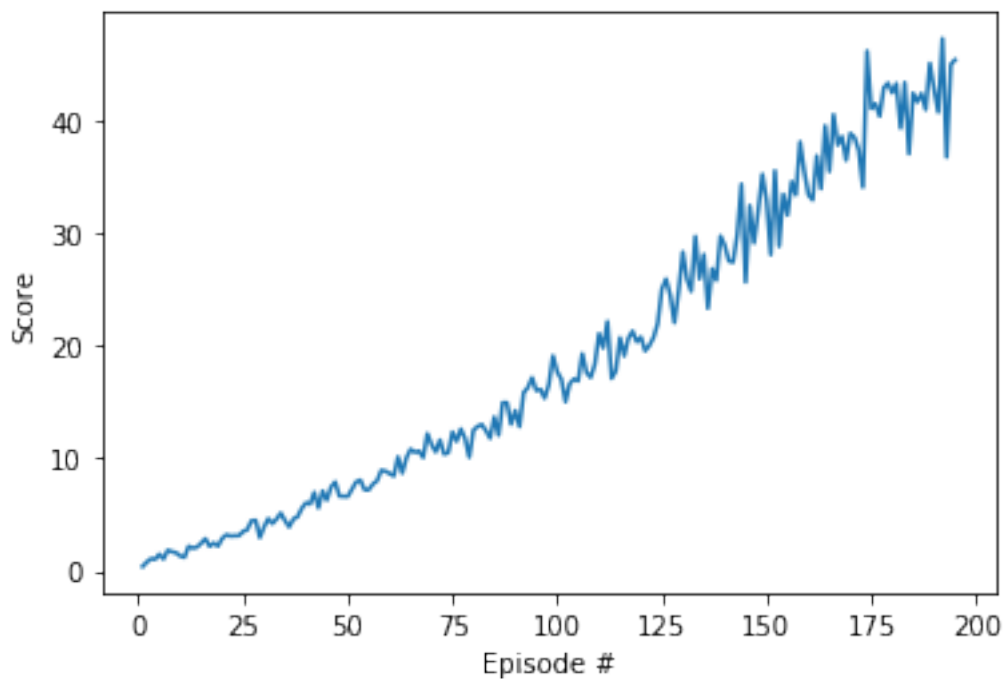


Figure 1: Learning

5 Ideas for making this better

Proximal Policy Optimization (PPO) and Distributed Distributional Deterministic Policy Gradients (D4PG) methods could be explored.

References

- [1] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.