Project criteria

The goal of this project was to train robotic arm to maintain contact with the sphere in the environment. The agents must get an average score of +30, over 100 consecutive episodes.

Implementation

The algorithm that I have used is an implementation of Deep Deterministic Policy Gradient with Replay Buffer. I have also found that features such as gradient clipping and learning every 10 episodes but 10 times helped to speed up training process.

Actor network:

- 3 layer neural network with additional batch normalization layer
- 33 -> 400 -> Batch norm -> 300 -> 4

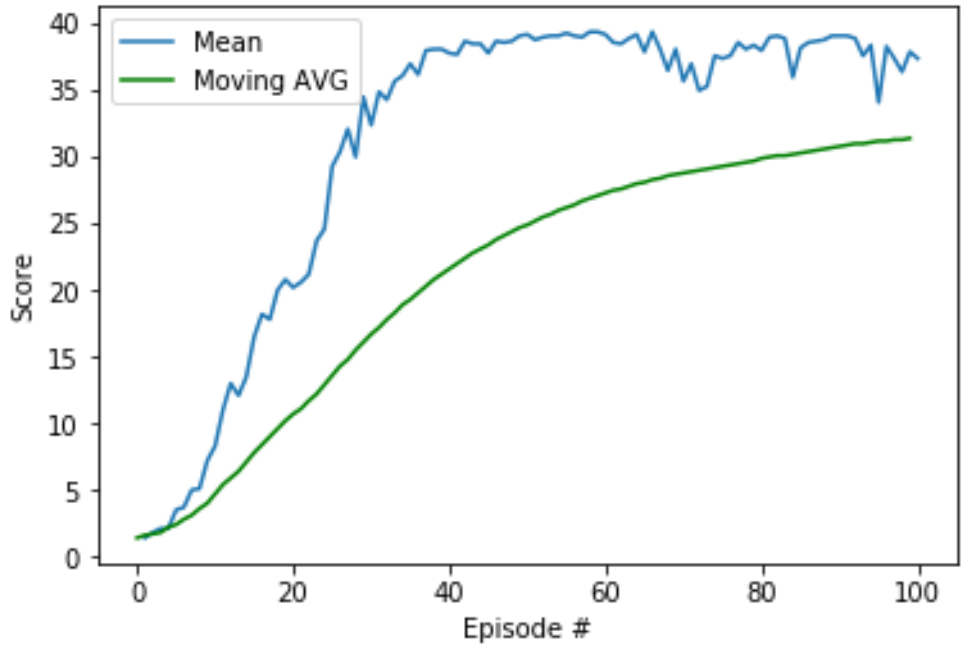
Critic network:

- 3 layer neural network with additional batch normalization layer
- 33 -> 400 -> Batch norm -> 300 -> 4

Hyperparameters:

- replay buffer size = 1e6
- minibatch size = 128
- discount factor = 0.99
- tau = 0.0025
- learning rate of the actor = 0.0002
- learning rate of the critic = 0.0005
- L2 weight decay = 0
- epsilon = 1
- epsilon decay = 1e-6
- epsilon min 0.1

Results



The agents were able to solve task in 100 episodes with a final moving average score of 31.4.

Improvements

- Experimenting with different RL algorithms may yield more stable training process
- Improving results tuning the hyperparameters
- Implement D3PG or D4PG