Project criteria

The goal of this project is is to train two seperate agents to play a game of tennis with each other. The agents must get a target score of +0.5, as the average, over 100 consecutive episodes.

Implementation

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

The actor-critic architecture:

Actor network:

- 3 layer neural network with additional batch normalization layer
- 24-> 400 -> Batch norm -> 200-> 2

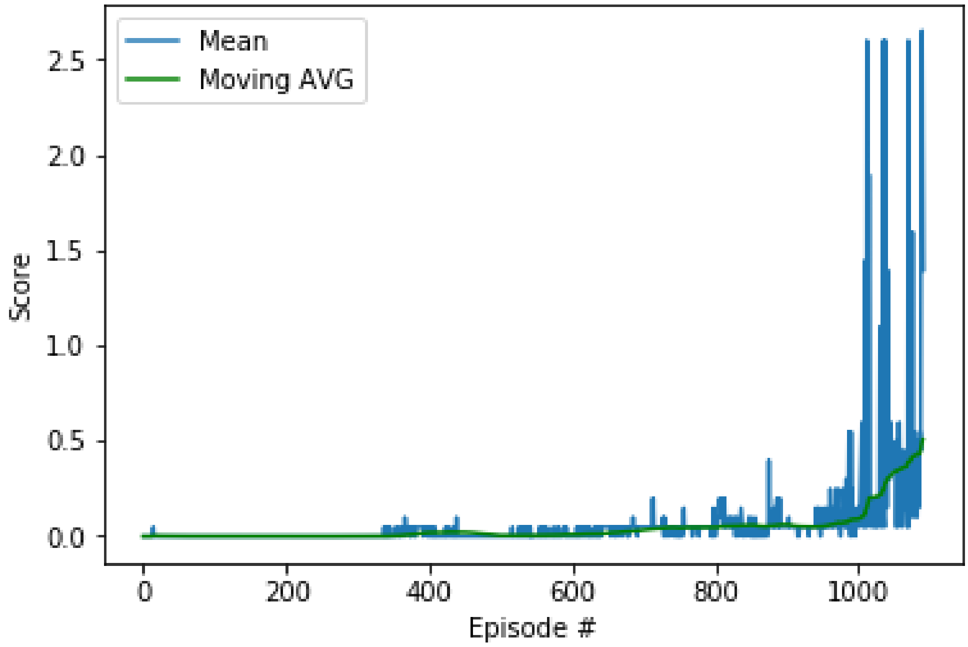
Critic network:

- 3 layer neural network with additional batch normalization layer
- 24-> 500-> Batch norm -> 300 -> 2

Hyperparameters:

- replay buffer size = 1e6
- minibatch size = 128
- discount factor = 0.99
- tau for soft update of target parameters = 0.0025
- learning rate of the actor = 0.0003
- learning rate of the critic = 0.0008
- L2 weight decay = 0
- epsilon = 1
- epsilon decay = 1e-6
- epsilon min 0.01

Results



The agents were able to solve task in 1091 episodes with a final moving average score of 0.5.

Improvements

- The competition and collaboration are the problems where the space for improvement is endless: from using different schedule for updating actor and critic networks to making different algorithms compete.
- Improving results tuning the hyperparameters
- Implement PPO, D3PG or D4PG