Project Report

Architechture

Actor

- Fully connected layer - state size to 400
- Fully connected layer - 400 to 300
- Fully connected layer - 300 to action size

Critic

- Fully connected layer - state size to 400
- Fully connected layer - 400+action size(since we concat) to 300
- Fully connected layer - 300 to 1

The following parameters were used:

buffer size: 1e6

- batch: 128
- gamma: 0.99
- tau: 1e-3
- learning rate for actor: 1e-4
- learning rate for critic: 1e-4
- weight decay: 0

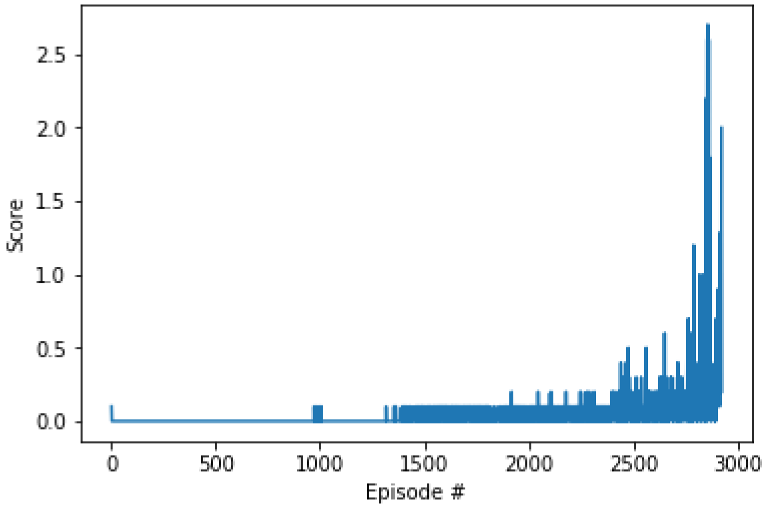
Learning algorithm

The learning algorithm which was implemented in this project is called Deep Determenistic Poilicy Gradient(DDPG) in its vanila form.

> *Deep Deterministic Policy Gradient (DDPG) is an algorithm which concurrently learns a Q-function and a policy. It uses off-policy data and the Bellman equation to learn the Q-function, and uses the Q-function to learn the policy.*

Results

Episode 2924 Average Score: 0.51 Environment solved in 2824 episodes! Average Score: 0.51



Future ideas

We could try to use L2 regularization on the weights, add some dropouts and also try to play more with hyperparameters and perform some gridsearch, antoerh option is to try other model such as AlphaZero as a baseline for this probelm.