UDACITY

DEEP REINFORCEMENT LEARNING

NANODEGREE UDACITY

Collaboration and Competition Project

1. Algorithm Details

Deep Deterministic Policy Gradient (DDPG) is a reinforcement learning algorithm designed for environments with continuous action spaces. It combines ideas from Q-learning and policy gradient methods and uses neural networks to approximate both the policy and the value function. Here are the key details and components of the DDPG algorithm

2. Network Architecture

A fully connected neural network with:
Input layer: 8 units
Hidden layers: Two hidden layers with 128 units each
Output layer: 2 units

3. Hyperparameters

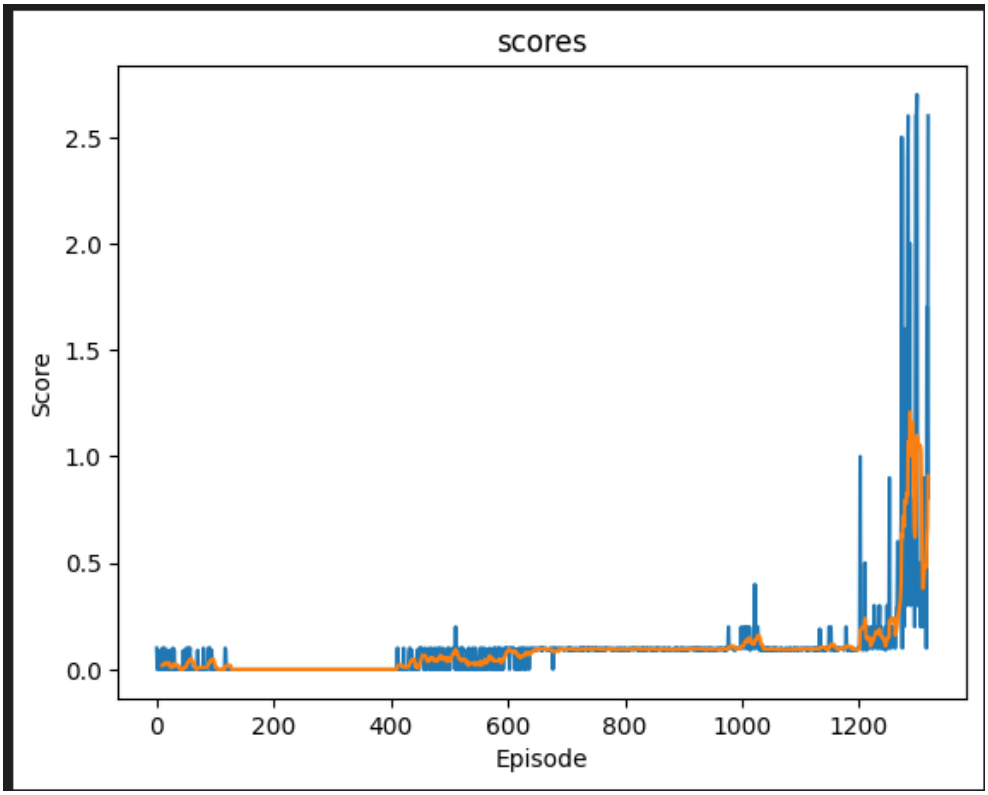
The following hyperparameters were used:

```
BUFFER_SIZE = int(1e5)  # replay buffer size
BATCH_SIZE = 128        # minibatch size
GAMMA = 0.99            # discount factor
TAU = 1e-3              # for soft update of target parameters
LR_ACTOR = 1e-4         # learning rate of the actor
LR_CRITIC = 1e-3        # learning rate of the critic
LEARN_EVERY = 1         # learn every LEARN_EVERY steps
LEARN_NB = 1            # how often to execute the learn-function every LEARN_EVERY steps
```

4. Plot of Rewards

I needed 1219 episodes to solve the environment:

```
Episode 100     Average Score: 0.02rage maximum score over the last 10 episodes: 0.03
Episode 200     Average Score: 0.00rage maximum score over the last 10 episodes: 0.00
Episode 300     Average Score: 0.00rage maximum score over the last 10 episodes: 0.00
Episode 400     Average Score: 0.00rage maximum score over the last 10 episodes: 0.00
Episode 500     Average Score: 0.04rage maximum score over the last 10 episodes: 0.05
Episode 600     Average Score: 0.05rage maximum score over the last 10 episodes: 0.09
Episode 700     Average Score: 0.09rage maximum score over the last 10 episodes: 0.10
Episode 800     Average Score: 0.10rage maximum score over the last 10 episodes: 0.10
Episode 900     Average Score: 0.10rage maximum score over the last 10 episodes: 0.10
Episode 1000    Average Score: 0.10rage maximum score over the last 10 episodes: 0.11
Episode 1100    Average Score: 0.11rage maximum score over the last 10 episodes: 0.10
Episode 1200    Average Score: 0.10rage maximum score over the last 10 episodes: 0.09
Episode 1300    Average Score: 0.44rage maximum score over the last 10 episodes: 1.10
Episode 1319    max score: 2.60 average maximum score over the last 10 episodes: 0.91
Environment solved in 1219 episodes!    Average Score: 0.52
```



5. Ideas for Future Work

To improve convergence speed, the developments covered in the D4PG course can be used to help reduce overestimation of action values