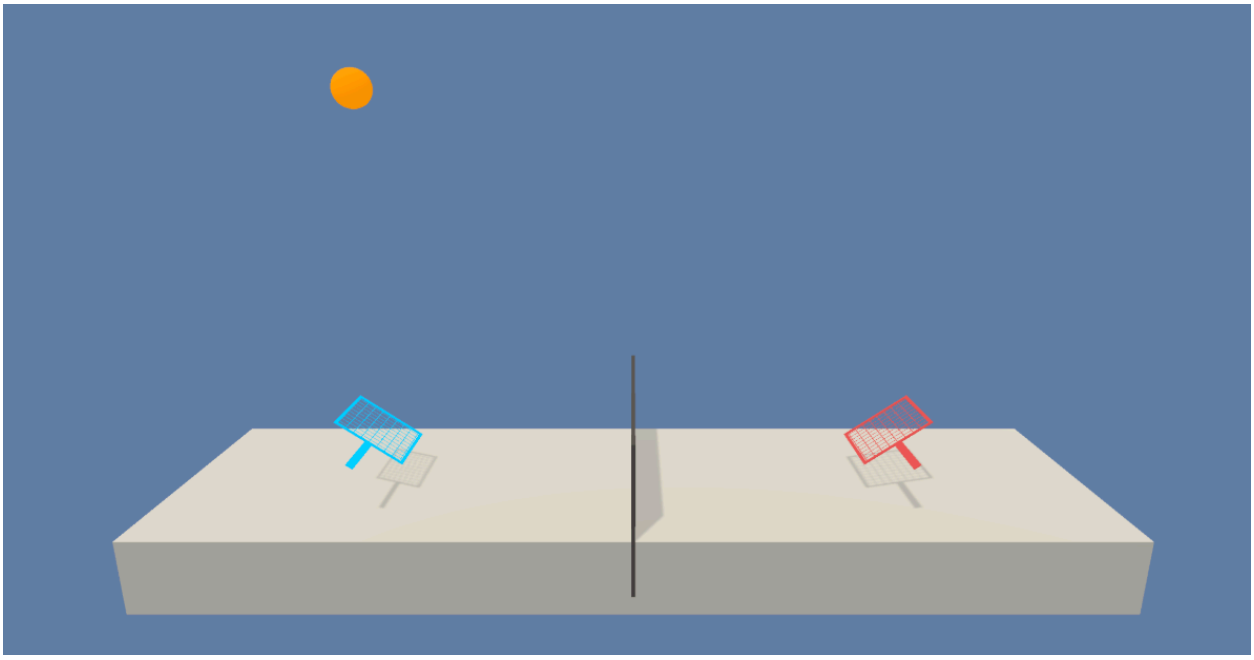
Ali Abdullah Al-Saeedi
Reinforcement Nanodegree
Continuous Control Project
April 5, 2019

ML-Agents Tennis



In this environment, two agents control rackets to bounce a ball over a net. If an agent hits the ball over the net, it receives a reward of +0.1. If an agent lets a ball hit the ground or hits the ball out of bounds, it receives a reward of -0.01. Thus, the goal of each agent is to keep the ball in play.

The observation space consists of 8 variables corresponding to the position and velocity of the ball and racket. Each agent receives its own, local observation. Two continuous actions are available, corresponding to movement toward (or away from) the net, and jumping.

1.Implementation

The solution is based on the MADDPG algorithm, using separated actor and critic for each agent and a shared memory buffer. The code is mostly the same as the project of continuous control, which was adapted from the DDPG-Pendulum exercise. The `Group` class, which is handling multiple agents, is capable of handling more than 2 agents.

Both actor and critic networks have 2 fully-connected hidden layers of **128 nodes**, both trained with a learning rate of **0.001**, using mini-batches of **256**, a replay buffer of **100000**, and a discount of **0.9**. The Ornstein-Uhlenbeck noise has a sigma of **0.1**and the soft update is made using a tau of **0.001**. The training seems to be fairly stable for a RL task. The score kept on improving after the goal was reached.

Hyperparameters:

There are many Hyperparameters that can be edited:

Number of episodes	n_episodes=10000
Max time per episodes	max_t=2000 (ms)

2.Results

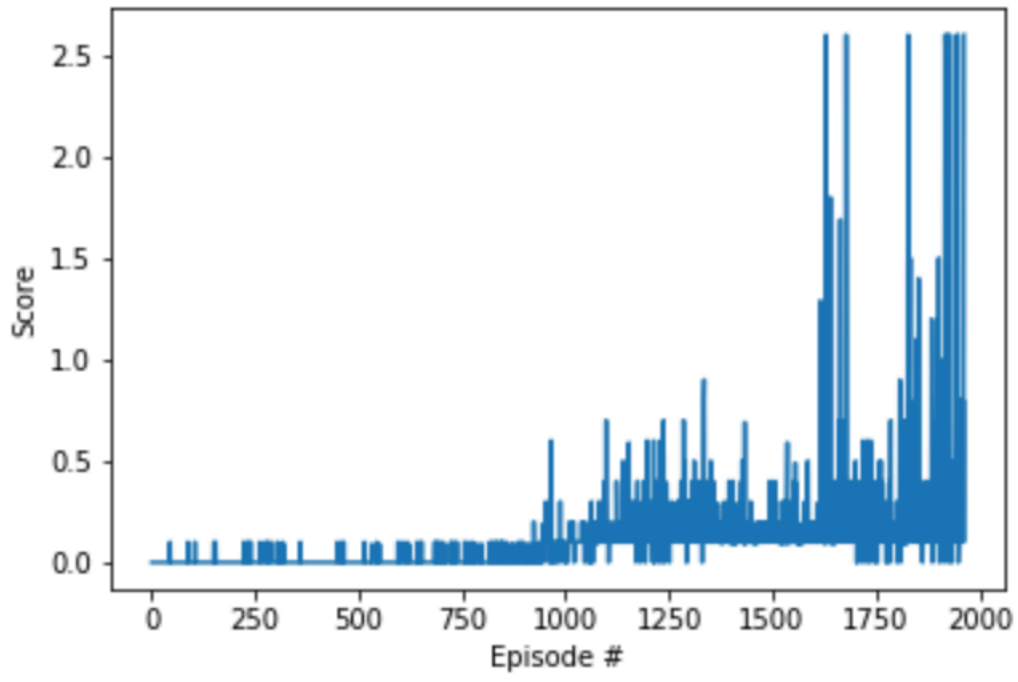
In the training we could solve the environment in 1964 episodes. By solving it we mean reaching an average score of at least 0.5.

```
In [7]: # Training our agent

        scores = group_ddpg(5000, 2000)
```

```
Episode 100     Average Score: 0.00
Episode 200     Average Score: 0.00
Episode 300     Average Score: 0.01
Episode 400     Average Score: 0.01
Episode 500     Average Score: 0.00
Episode 600     Average Score: 0.00
Episode 700     Average Score: 0.01
Episode 800     Average Score: 0.02
Episode 900     Average Score: 0.03
Episode 1000    Average Score: 0.06
Episode 1100    Average Score: 0.12
Episode 1200    Average Score: 0.16
Episode 1300    Average Score: 0.20
Episode 1400    Average Score: 0.20
Episode 1500    Average Score: 0.17
Episode 1600    Average Score: 0.18
Episode 1700    Average Score: 0.32
Episode 1800    Average Score: 0.21
Episode 1900    Average Score: 0.33
Episode 1964    Average Score: 0.52
Environment solved in 1964 episodes!    Average Score: 0.52
```

This graph shows how while our agent is training we are gaining a better result.



3. Future Work

- Try different algorithms than the DDPG for each of the agents
- Try to fix stabilization for the multi-agent systems.