

# Project Report

## Learning algorithm

The learning algorithm used is Multi Agent Deep Deterministic Policy Gradient Q Learning as presented in this paper <https://arxiv.org/pdf/1706.02275.pdf>.

### Agent 1:

**The actor network has following layers:**

- Fully connected layer - input: 24 (state size) output: 300
- Fully connected layer - input: 300 output 200
- Fully connected layer - input: 200 output 64
- Fully connected layer - input: 64 output: 2 (action size)

**The critic network has following layers:**

- Fully connected layer - input: 52 (state size+action\_size)\*2, output: 400
- Fully connected layer - input: 400 output 300
- Fully connected layer - input: 300 output: 1

**Parameters used for the DDPG agent:**

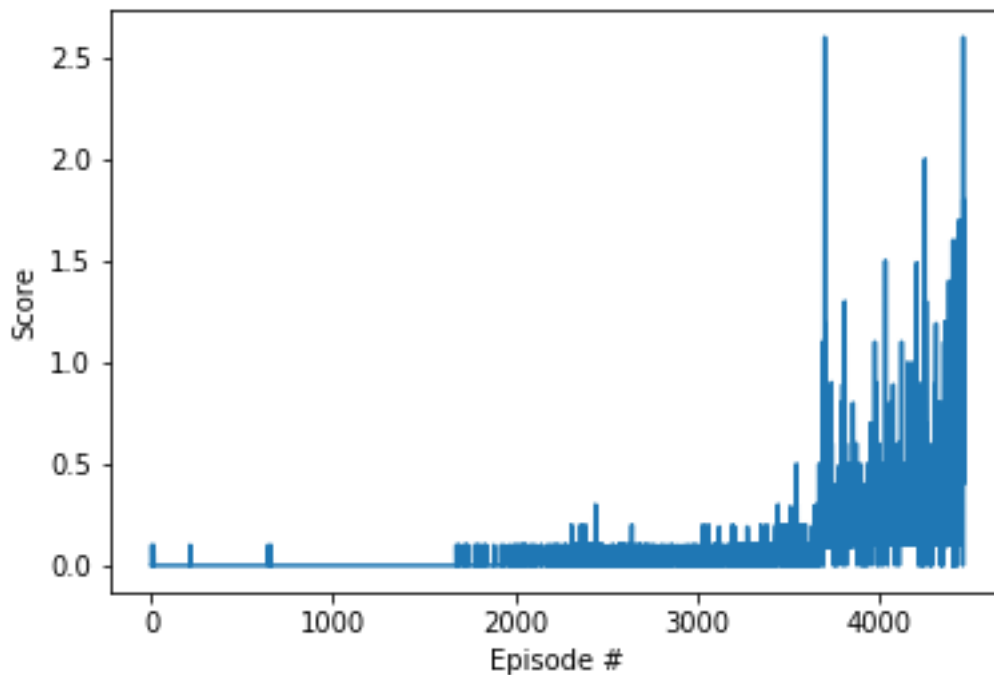
- BUFFER\_SIZE = 2000000 # replay buffer size
- BATCH\_SIZE = 200 # minibatch size
- GAMMA = 0.99 # discount factor
- TAU = 1e-3 # for soft update of target parameters
- LR\_ACTOR = 0.0001 # learning rate of the actor
- LR\_CRITIC = 0.001 # learning rate of the critic
- WEIGHT\_DECAY = 0.00001 # L2 weight decay

## Results :

Episode 0	Average Score: 0.000
Episode 100	Average Score: 0.002
Episode 200	Average Score: 0.000
Episode 300	Average Score: 0.001
Episode 400	Average Score: 0.000
Episode 500	Average Score: 0.000
Episode 600	Average Score: 0.000
Episode 700	Average Score: 0.002
Episode 800	Average Score: 0.000
Episode 900	Average Score: 0.000
Episode 1000	Average Score: 0.000
Episode 1600	Average Score: 0.000
Episode 1700	Average Score: 0.004
Episode 1800	Average Score: 0.008
Episode 1900	Average Score: 0.005
Episode 2000	Average Score: 0.004

Episode 2100	Average Score: 0.016
Episode 2200	Average Score: 0.009
Episode 2300	Average Score: 0.010
Episode 2400	Average Score: 0.044
Episode 2500	Average Score: 0.061
Episode 2600	Average Score: 0.045
Episode 2700	Average Score: 0.033
Episode 2800	Average Score: 0.020
Episode 2900	Average Score: 0.015
Episode 3000	Average Score: 0.015
Episode 3100	Average Score: 0.022
Episode 3200	Average Score: 0.022
Episode 3300	Average Score: 0.050
Episode 3400	Average Score: 0.070
Episode 3500	Average Score: 0.088
Episode 3600	Average Score: 0.092
Episode 3700	Average Score: 0.119
Episode 3800	Average Score: 0.209
Episode 3900	Average Score: 0.221
Episode 4000	Average Score: 0.224
Episode 4100	Average Score: 0.218
Episode 4200	Average Score: 0.274
Episode 4300	Average Score: 0.336
Episode 4400	Average Score: 0.378

Environment solved in 4469 episodes! Average Score: 0.502



### Future work:

- Better hyperparameter tuning
- Try different architectures for the models
- Try other algorithms like REINFORCE, TNPG, RWR, REPS, TRPO, CEM, CMA-ES and compare them to DDPG