

# Project Report

## Learning Algorithm

**The following algorithms was used as part of the solution:**

1 – Continuous Control with Deep Reinforcement Learning (**DDPG**):

<https://arxiv.org/pdf/1509.02971.pdf>

2 - Dueling Network Architectures: <https://arxiv.org/pdf/1511.06581.pdf>

**The architecture used was:**

Agent 1:

```
Actor Local and Actor Target (  
    (fc1): Linear(in_features=24, out_features=400, bias=True)  
    (fc2): Linear(in_features=400, out_features=300, bias=True)  
    (fc3): Linear(in_features=300, out_features=2, bias=True)  
)  
Critic Local and Critic Target(  
    (fc1): Linear(in_features=52, out_features=400, bias=True)  
    (fc2): Linear(in_features=400, out_features=300, bias=True)  
    (fc3): Linear(in_features=300, out_features=1, bias=True)  
)
```

Agent 2 (twin agent):

```
Actor Local and Actor Target(  
    (fc1): Linear(in_features=24, out_features=550, bias=True)  
    (fc2): Linear(in_features=550, out_features=300, bias=True)  
    (fc3): Linear(in_features=300, out_features=2, bias=True)  
)  
Critic Local and Critic Target(  
    (fcs1): Linear(in_features=52, out_features=550, bias=True)  
    (fc2): Linear(in_features=550, out_features=300, bias=True)  
    (fc3): Linear(in_features=300, out_features=1, bias=True)  
)
```

## Hyperparameters

BUFFER\_SIZE = 50000

BATCH\_SIZE = 128

GAMMA = 0.995

TAU = 0.002

LR\_ACTOR = 0.0001

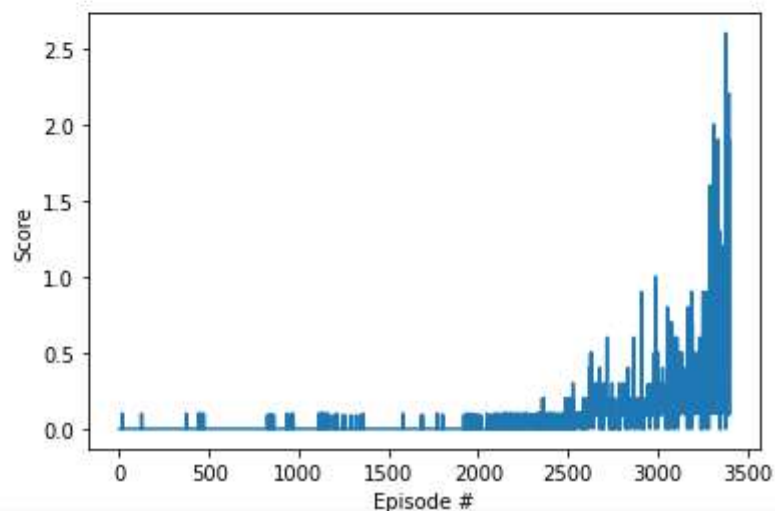
LR\_CRITIC = 0.001

UPDATE\_EVERY = 2

## Plot of Rewards

Episode 0	Average Score: 0.000
Episode 200	Average Score: 0.001
Episode 400	Average Score: 0.001
Episode 600	Average Score: 0.000
Episode 800	Average Score: 0.000
Episode 1000	Average Score: 0.004
Episode 1200	Average Score: 0.017
Episode 1400	Average Score: 0.003
Episode 1600	Average Score: 0.001
Episode 1800	Average Score: 0.002
Episode 2000	Average Score: 0.014
Episode 2200	Average Score: 0.025
Episode 2400	Average Score: 0.057
Episode 2600	Average Score: 0.083
Episode 2800	Average Score: 0.099
Episode 3000	Average Score: 0.133
Episode 3200	Average Score: 0.197
Episode 3400	Average Score: 0.490

Environment solved in 3301 episodes!



## Ideas for future work

- 1 – Adapt this implementation to use only one critic network for both agents.
- 2 – Implement "Prioritize Experience Replay", instead a ReplayBuffer.