Udacity Deep Reinforcement Learning Project 3: Collaboration and Competition

For training Deep Deterministic Policy Gradients (DDPG) is used as described in the provided paper: https://arxiv.org/pdf/1509.02971.pdf

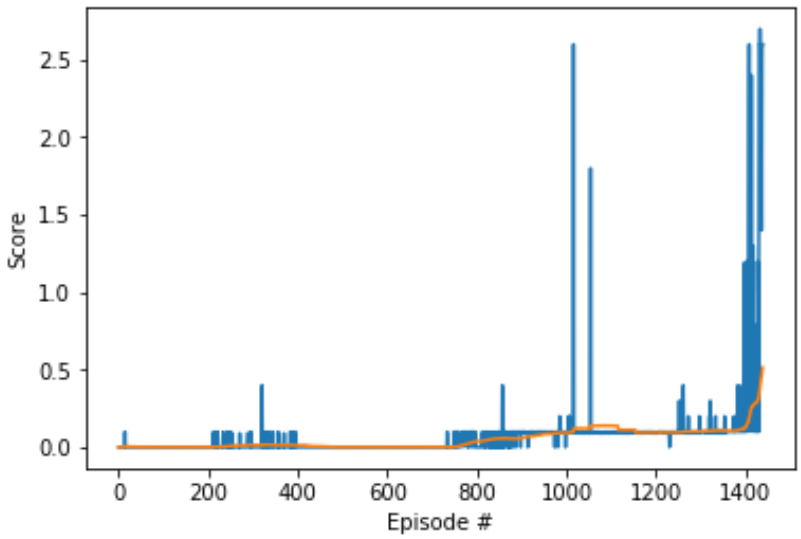
DDPG uses an actor-critic architecture to handle large action spaces as the one in the given exercise. It can only be used for policy based problems. It deals with the policy function independently of the value function. The implemented algorithm uses experience replay to stabilize the training. The algorithm is modified to handle actions for multiple agents with a shared memory.

For actor and critic separate neural networks are used. Both networks consist of three fully connected layers with ReLu activation. Batch normalization is applied to the first layer. The two hidden layers of each network have a size of 128 nodes. The actor network has 24 inputs and 2 outputs. The critic network has 2 inputs and 1 output.

The following hyperparameters are used:

```
BUFFER_SIZE = int(1e5)  # replay buffer size
BATCH_SIZE = 128        # minibatch size
GAMMA = 0.99            # discount factor
TAU = 0.001             # for soft update of target parameters
LR_ACTOR = 2e-4         # learning rate of the actor
LR_CRITIC = 2e-4        # learning rate of the critic
WEIGHT_DECAY = 0        # L2 weight decay
```

These give quite good results after a few episodes:




```
Problem Solved in 1438 episodes.    Average Score: 0.5142
```

In future works other mtehodes like Proximal Policy Optimization could be explored.