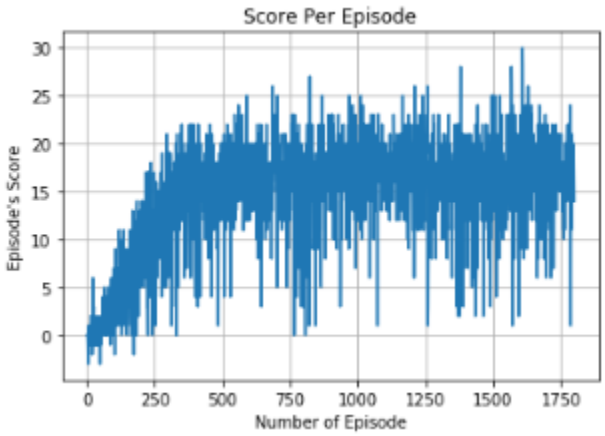
UDACITY DEEP REINFORCMENT LEARNING – PROJECT 1 REPORT "BANANA RL"

Mahmoud Adel Nour

m.adel.nour@gmail.com

Report

1. Learning Algorithm
    - Building the Deep Q-Learning network using 2 hidden layers with each 64 neurons and the optimizer is ADAM
    - Epsilon-greedy action selection is implemented to select random action but it is set to 0% by default and it didn't give better performance over the current implementation
    - Using Experience replay and review every 64 random replays and check every 4 record if we reached the 64 replays to do the learning
    - Creating target weights and current(local) weights knowing that the target weights changes with small steps from the current weights to keep soft update concept
    - The current implementation achieved mean score over 100 episodes greater than 13 after 400 episodes of training
    - Hyper parameters:
        - BUFFER_SIZE = 100000 : replay buffer size
        - BATCH_SIZE = 64 : batch size to start learning over the stored replays
        - GAMMA = 0.99 : discount factor
        - TAU = 0.001 : soft update factor which control the change of the weights
        - LR = 0.0005 : learning rate for the optimizer "ADAM"
2. Plot of Rewards



3. Ideas for Future Work
    - to use more complicated algorithm like Rainbow DQN or Noisy DQN