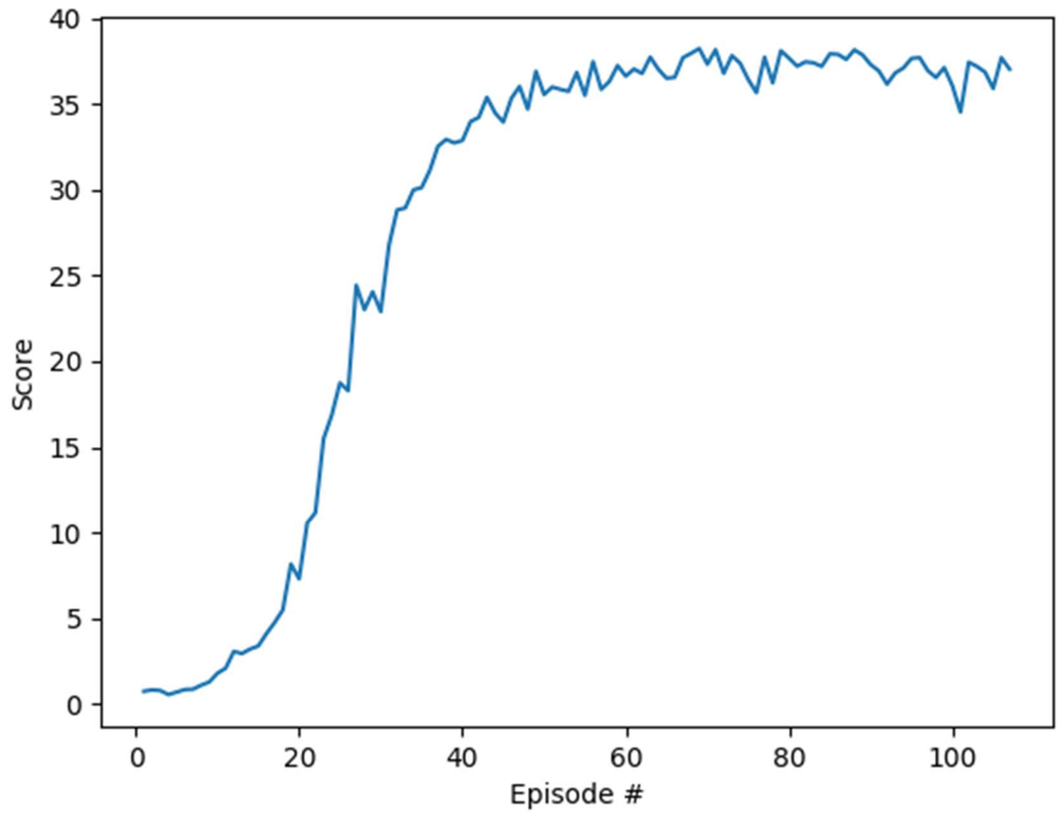
Project 2: Continuous Control with Unity Reacher Environment

Algorithm description

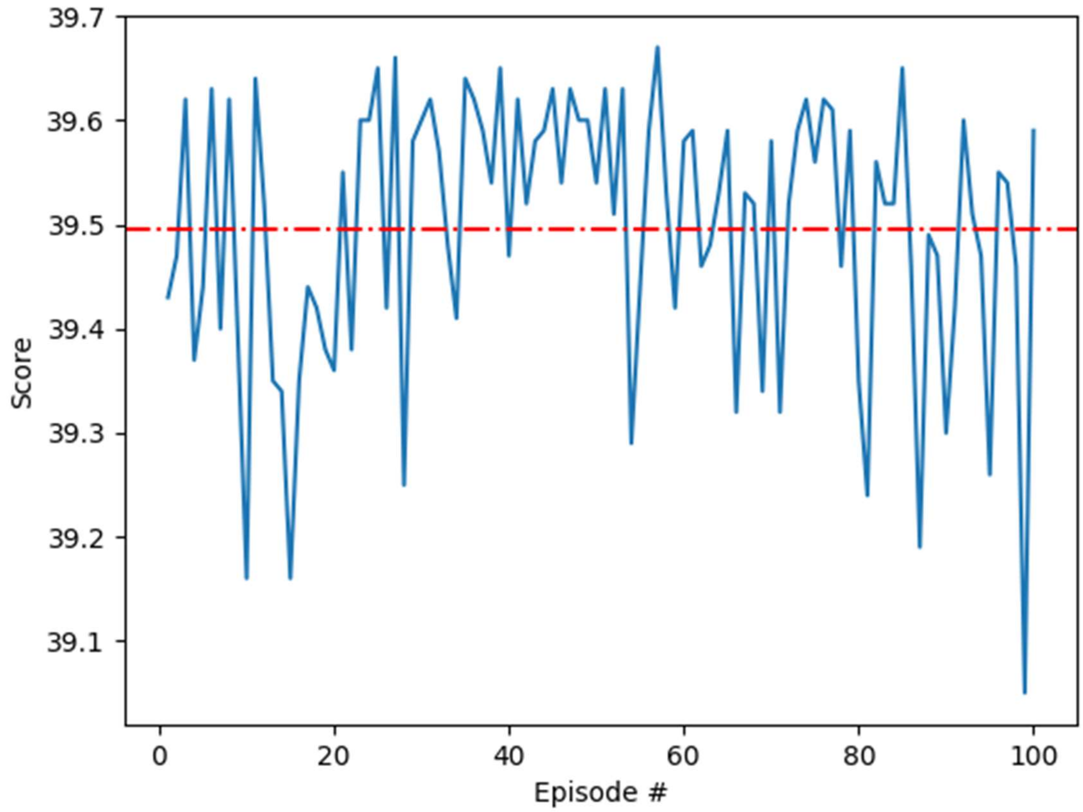
The Deep Deterministic Policy Gradient (DDPG) algorithm was used to solve this environment, specifically the version of the environment designed to work with multiple agents. DDPG is an actor-critic method that entails an actor to produce optimal actions according to a policy, and a critic to predict the state-action value of following the policy. It derives from Deep Q-Network (DQN) and Deterministic Policy Gradient (DPG), enabling the jump to continuous action spaces [1]. Most of this code was provided in the Udacity nanodegree repository, and changes were made to the replay buffer and the noise model to accommodate multiple agents.

Training and Testing Results

The environment was solved by the DQN agent in 107 episodes, obtaining an average score over the previous 100 episodes of +30. Learning is initially slow; but once improvement begins, the agent sees substantial gains for a period. It then settles into a more gradual learning.



The trained agent was run for an additional 100 episodes to prove its capability to generalize to new states. These results are included here.



Opportunities for Future Work

With further time, it is likely that the hyperparameters could be optimized to produce quicker learning. The current hyperparameters produce very consistent testing results using a trained agent.

References

[1] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra, “Continuous control with deep reinforcement learning”, https://arxiv.org/abs/1509.02971