**UDACITY DRLND**

**P2-CONTINUOUS-CONTROL**

**Description :**

The challenge requires to train a double jointed arm agent to move to target locations. A reward of +0.1 is provided for each step that the agent's hand is in the goal location and the agent needs to maintain its position at target.

---

**Learning Algorithm : DDPG**

The learning algorithm is Deep Deterministic Policy Gradient which is used to solve the Bipedal Gym environment in the session. The agent is unaware of the entire state and tries to get reward through following policy based algorithm. The architecture employed to solve this environment is as follows:

Actor Network:

2 Fully connected layers with ReLU activations and tanh on the last layer.  Learning Rate: 1e-4

Critic Network:

4 Fully connected layers with  leaky ReLU activations. Learning Rate: 3e-4

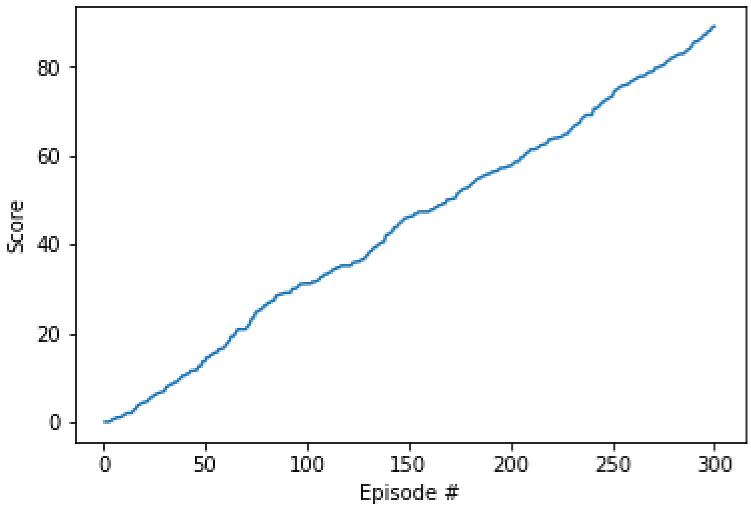
Parameters:

Replay buffer size : 1000000, Minibatch size:128,  Tau = 1e-3

Discount factor: 0.99, Weight decay: 0.0001

*Please note that the deadline is Jan 15th

**Agent and Environment:** The environment is solved in 149 episodes.



**Future Work:**

DDPG algorithm works quite well in this environment. Further, hyperparameter tuning would have given improvements given the reacher environment is solved in few hundred episodes, we can confidently try PPO, TRPO and actor critic methods such as A3Cs. I will follow these github repositories to improve upon my knowledge and to write modular code. :)

*https://github.com/qfettes/DeepRL-Tutorials*

*https://github.com/ShangtongZhang*

As always I will watch out for recent trends from OPENAI blog.

*Please note that the deadline is Jan 15th