**Project 2 Report – DDPG**

**1 Learning Algorithm**

In this project we are using a model-free, off-policy actor-critic algorithm using neural networks as deep function approximators for high dimensional policies with continuous action spaces.

The way actor-critic works is that we set the actor as the approximator for the policy which selects action and the critics that evaluate policy according to the action taken. The algorithm we are using in this project is known as Deep Deterministic Gradient Policy (DDPG), which is like DQN but for continuous spaces, as opposed to discrete ones.

It is important to highlight there are three specific advances made to DQN which will be applied here to make Deep networks work and perform in a stable manner, namely the following:

1. We minimise correlation between samples - by incorporating training from samples taken from batches of a replay buffer trained off-policy
2. TD backups synchronisation – we perform updates to target values by using a trained Q-network under the temporal difference backups to achieve consistency through training.
3. Batch normalisation – tackles the problem of a phenomena known as internal covariate shift which brings up training time due to the reduction in the learning rate because of the increasing number of nonlinearities. To counter act this problem normalisation is implemented to each training mini-batch.

**1.1 Model Architectures**

**1.1.1 Actor Network**

For the Actor neural network

**1.1.2 Critic Network**

**1.2 Hyperparameters**

**2 Plot of Rewards**

**3 Ideas for Future Work**