Project Report

Deep Deterministic Policy Gradients Algorithm has been used train the agents to play tennis. As the action space is continuous, I decided to use the Deep Deterministic Policy Gradients Algorithm.

I have used **two neural networks**, the actor network and the critic network.

The **actor network** contains two hidden layers of 128 units and 64 units respectively.
Relu activation has been applied on both the layers.
Tanh activation is applied on the final fully connected layer.
Batch normalization is also applied on the input to actor network.

Critic network also has two hidden layers of sizes 128 units and 64 units with relu activation on both the layers.
No activation is applied on the final fully connected layer.
Batch normalization is also applied on the input to critic network.

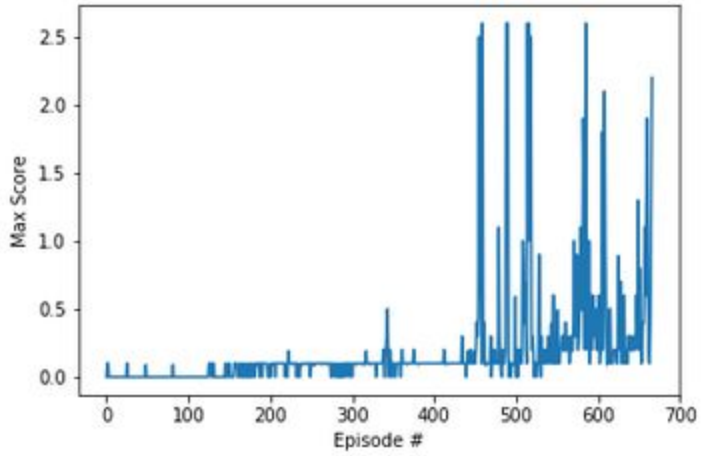
A lot of tuning and tweaking was done with the **hyperparameters**. Finally, I ended up using the following hyperparameters.

1. Buffer Size - 1e-6
2. Batch Size - 512
3. Gamma - 0.99
4. Tau - 1e-3
5. Learning Rates - 5e-4 (Actor)
6. Learning Rates - 1e-3 (Critic)

Plot of rewards

- Environment is solve in **667 episodes**.

```
Average Score: 0.51
```



Work possible in future

- Changing the actor and critic network sizes in DDPG.
- Trying algorithms other than DDPG such as PPO, A3C, D4PG.