Learning algorithm

Owing to the continuous action space, Deep Deterministic Policy Gradients algorithm has been used for this project.

Network acrhitectures

Actor:

- Hidden layer - 128 units
- Hidden layer - 64 units
- Batch norm and relu activation

Critic:

- Hidden layer - 256 units
- Hidden layer - 64 units
- Hidden layer - 32 units
- Relu activation and batch norm on input

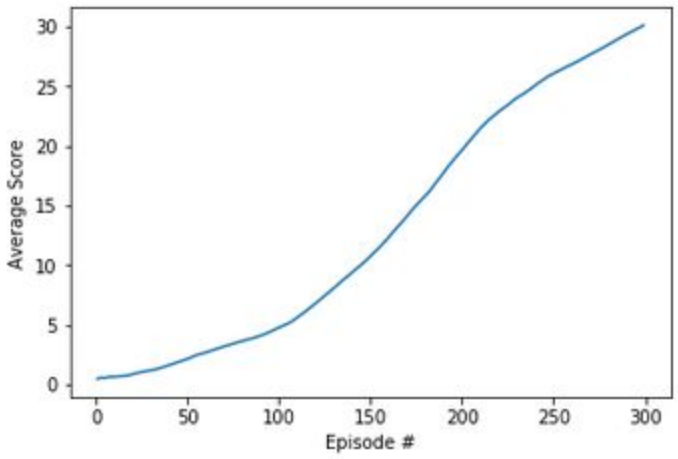
Hyperparameters

Following hyperparameters were used to train the agent.

```
BUFFER_SIZE = int(1e6)  # replay buffer size
BATCH_SIZE = 512        # minibatch size
GAMMA = 0.99            # discount factor
TAU = 1e-3              # for soft update of target parameters
LR_ACTOR = 9e-4         # learning rate of the actor
LR_CRITIC = 1e-3        # learning rate of the critic
WEIGHT_DECAY = 0.0001   # L2 weight decay
```

Reward plot

At around episode 300 average score of 30+ was obtained.



Ideas to improve the performance of the agent

- I experimented only with DDPG. Other algorithms also need to be tried in future to gain improvement in performance like PPO, A3C etc.