Learning Algorithm:

The learning algorithms implemented were DQN (Deep Q-Learning Algorithm with Neural Network) with attempted improvements through Double DQN (DDQN), Prioritized Experienced Replay, Dueling DQN all with Fixed-Q targets. The state has 37 dimensions and with possible 4 actions. Two identical networks were created, local and target, where target model provides Q values based on next states, and local model provides expected Q values based on current states. The goal of the neural network learning process is to minimize the error between these two values.

All the algorithms use a fixed buffer of 10,000 size, and was trained over 1,000 episode with each episode having 1,000 steps. The model was trained every 4 steps, and the experience from each step was added to the fixed buffer (which deleted oldest experience when full to accommodate for newest experiences).

A Deep Neural Network was used to model the Q-table
For DQN, DDQN:
    37 inputs for each dimension→ ReLu activation → 64 unit hidden layer → ReLu activation → 64 unit hidden layer → 4 unit output for each action
For Dueling DQN:
    Feature = 37 inputs for each dimension→ ReLu activation → 64 unit hidden layer → ReLu activation → 64 unit hidden layer
    Value = Output of Feature → ReLu activation → 1 unit output
    Advantage = Output of Feature → ReLu activation → 4 unit output for each action
    Final Q = Value + Advantage

For regular Experience Replay, 64 experiences were sampled randomly for the buffer, whereas for Prioritized Experience Replay, 64 experiences were sampled using a weight proportional to the loss of the sample between target and local models, with the sampling weights updated whenever the particular experience was sampled.

Other hyperparameters (applicable to all algorithms)
    GAMMA = 0.99 # discount factor
    TAU = 1e-3 # for soft update of target parameters
    LR = 5e-4 # learning rate

Prioritized replay buffer hyperparameters
    e = 0.01
    a = 0.6
    beta = 0.4
    beta_increment_per_sampling = 0.001

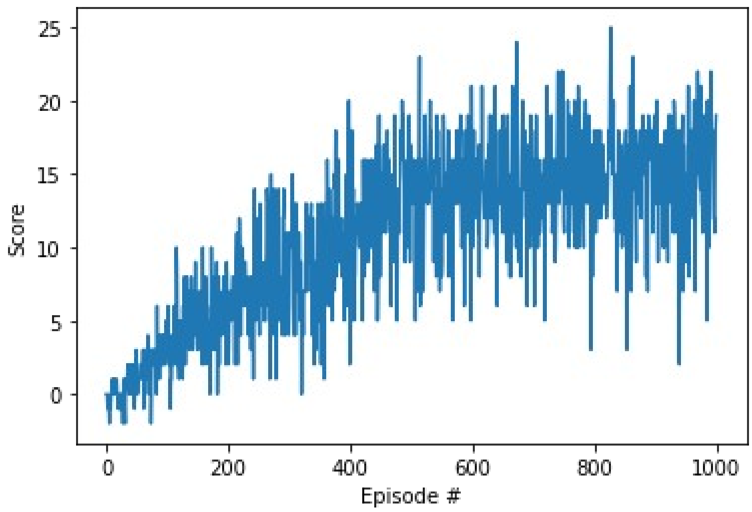
Various hyperparameters were tried by changing the hyperparameters, but best results (below) were achieved with the hyperparameter values above. Best algorithm was Dueling DDQN, based on the scores below

Plot of Rewards

<u>DQN</u>

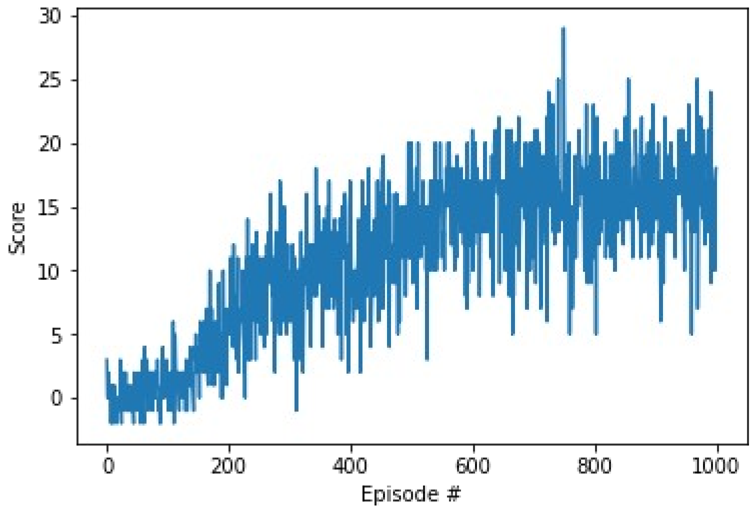
```
Episode 100	Average Score: 1.14
Episode 200	Average Score: 4.82
Episode 300	Average Score: 7.48
Episode 400	Average Score: 9.45
```

```
Episode 500     Average Score: 12.82
Episode 513     Average Score: 13.00
Environment solved in 413 episodes!     Average Score: 13.00
Episode 600     Average Score: 13.90
Episode 700     Average Score: 14.33
Episode 800     Average Score: 14.49
Episode 900     Average Score: 15.05
Episode 977     New Best Average Score: 15.21
Episode 1000    Average Score: 15.00
```



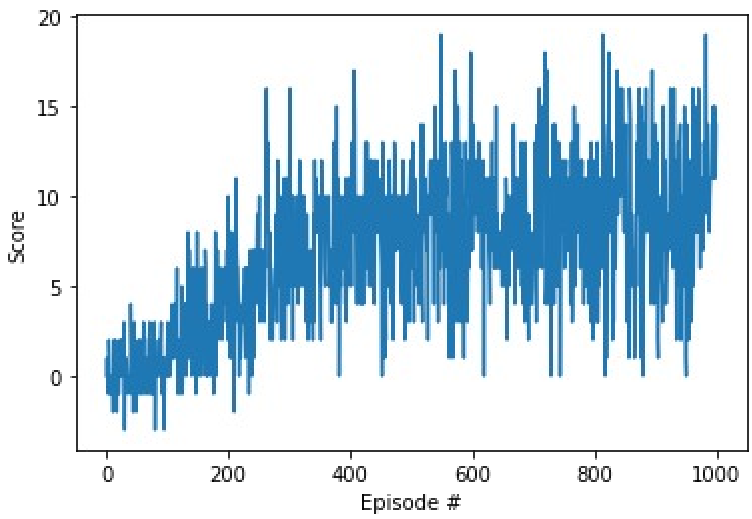
Double DQN (DDQN)

```
Episode 100     Average Score: 0.41
Episode 200     Average Score: 2.66
Episode 300     Average Score: 7.80
Episode 400     Average Score: 9.77
Episode 500     Average Score: 11.52
Episode 546     Average Score: 13.01
Environment solved in 446 episodes!     Average Score: 13.01
Episode 600     Average Score: 14.31
Episode 700     Average Score: 15.22
Episode 800     Average Score: 15.78
Episode 900     Average Score: 16.03
Episode 991     New Best Average Score: 16.40
Episode 1000    Average Score: 16.29
```



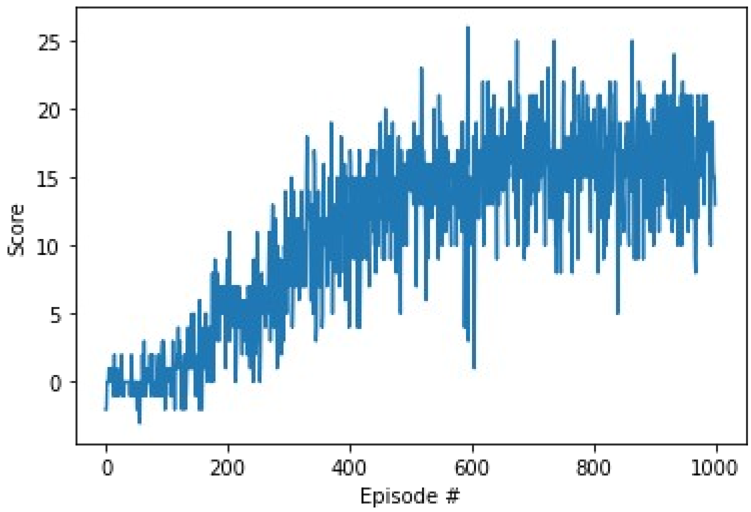
Prioritized DDQN

```
Episode 100     Average Score: 0.30
Episode 200     Average Score: 2.73
Episode 300     Average Score: 5.50
Episode 400     Average Score: 7.35
Episode 500     Average Score: 8.00
Episode 600     Average Score: 8.86
Episode 700     Average Score: 8.38
Episode 800     Average Score: 9.15
Episode 900     Average Score: 9.31
Episode 1000    Average Score: 9.42
```



Dueling DDQN

```
Episode 100     Average Score: 0.03
Episode 200     Average Score: 2.30
Episode 300     Average Score: 5.74
Episode 400     Average Score: 10.44
Episode 500     Average Score: 13.01
Environment solved in 400 episodes!     Average Score: 13.01
Episode 600     Average Score: 14.49
Episode 700     Average Score: 15.86
Episode 800     Average Score: 16.26
Episode 900     Average Score: 15.97
Episode 963     New Best Average Score: 16.73
Episode 1000    Average Score: 16.49
```



Ideas for Future Work

While best results were obtained from the Dueling DDQN, other algorithms such as Learning from multi-step bootstrap targets, Distributional DQN, and Noisy DQN can be attempted, with combinations of algorithms to further improve the results.