Project report

1. Implementation of learning algorithm

The first learning algorithm I implanted is vanilla Deep Q Learning with fixed Q-target and random experience replay by making only minor modifications to the DQN code provided as part of the Deep Q-Networks lesson. This allows me to familiar with the environment and get my code working with the Unity Machine Learning Agents.

Then I tried implementing Dueling DQN where output of 3rd fully connected layer splits in two different streams i.e. one stream for value function V(s) and the other stream for advantage function A(s, a). Finally, the Q-values are the combined result of the value V and the advantage A for each action.

DQN architecture

Input (state size: 37)
Fully connected layer – 64 nodes
Relu
Fully connected layer – 64 nodes
Relu
Fully connected layer – 64 nodes
Output (Q values)

Dueling DQN architecture

Input (37 states)	
Fully connected layer – 64 nodes	
Relu	
Fully connected layer – 64 nodes	
Relu	
Fully connected layer – 64 nodes	
Relu	
Value Stream: Fully connected layer – 128 nodes	Advantage Stream: Fully connected layer – 128 nodes
Relu	Relu
Fully connected layer – 1 node	Fully connected layer – 4 nodes
Output (Q values)	

Hyperparameters:

- Replay buffer size = 1e5
- Batch size = 128
- Discount factor (GAMMA) = 0.99
- TAU (for soft update of target parameters) = 1e-3
- Learning rate = 5e-4
- Update network frequency: every 4 steps

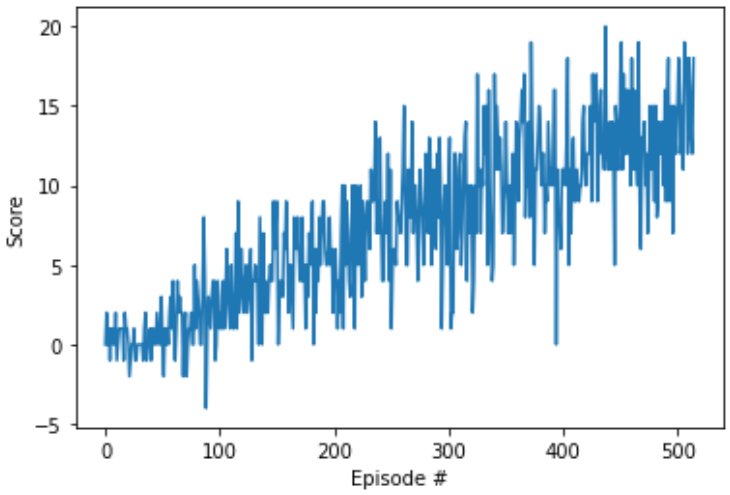
- Maximum steps per episode: 2000
- Maximum episodes = 2000
- Starting epsilon = 1.0
- Ending epsilon = 0.01
- Epsilon decay rate = 0.995

2. Results

DQN

Training Output:

```
Episode 100     Average Score: 0.95
Episode 200     Average Score: 4.46
Episode 300     Average Score: 7.67
Episode 400     Average Score: 10.23
Episode 500     Average Score: 12.33
Episode 515     Average Score: 13.04
Environment solved in 415 episodes!         Average Score: 13.04
```

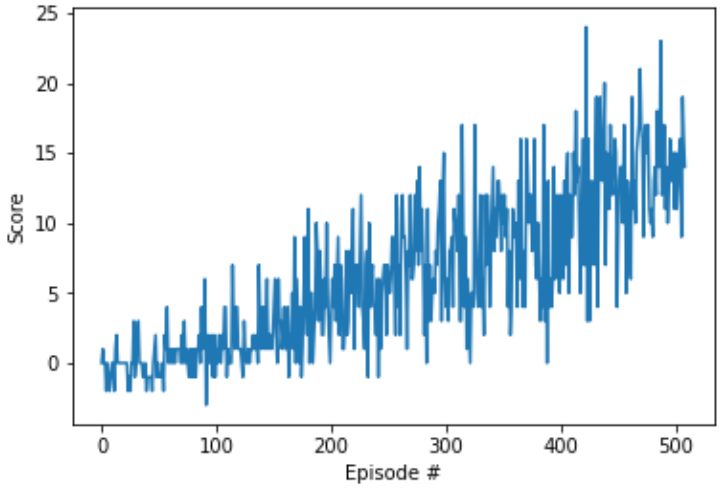


Dueling DQN

Training Output:

```
Episode 100     Average Score: 0.23
Episode 200     Average Score: 2.67
Episode 300     Average Score: 6.31
Episode 400     Average Score: 8.15
Episode 500     Average Score: 12.67
Episode 509     Average Score: 13.01
Environment solved in 409 episodes!         Average Score: 13.01
```

Training Plot:



3. Ideas for future work

1. Double Deep Q Networks
2. Prioritized Experience Replay
3. Rainbow: Combining Improvements in Deep Reinforcement Learning