Project Report

Learning Algorithm

- **Fixed Q-Targets**: Two networks are used, with the “target” network merely being a copy of the training network. The training network will have its Q-values updated, but will use the Q-value of the target network when estimating future discounted return (i.e., it will use the “old” network to generate estimates). The target network will periodically be updated to match the training network, so that these estimates won’t be far away from the training network. Decoupling the target from the parameters makes the learning algorithm much more stable and less likely to diverge or fall into oscillations.
- **Hyperparameters**: The same hyperparameters of the agent and the training were obtained from DQN project source code in Udacity workspace.
  - **Agent Parameters:**

Parameter	Value
replay buffer size	100000
Minibatch size	64
Discount factor	0.99
Tau for soft update of target parameters	0.001
Learning rate	0.0005
How often to update the network	Update every 4 steps

  - **Training Parameters:**

Parameter	Value
Number of episodes	1000
Maximum iteration	1000
Epsilon start	1.0
Epsilon End	0.01
Epsilon Decay	0.995

Model Architecture

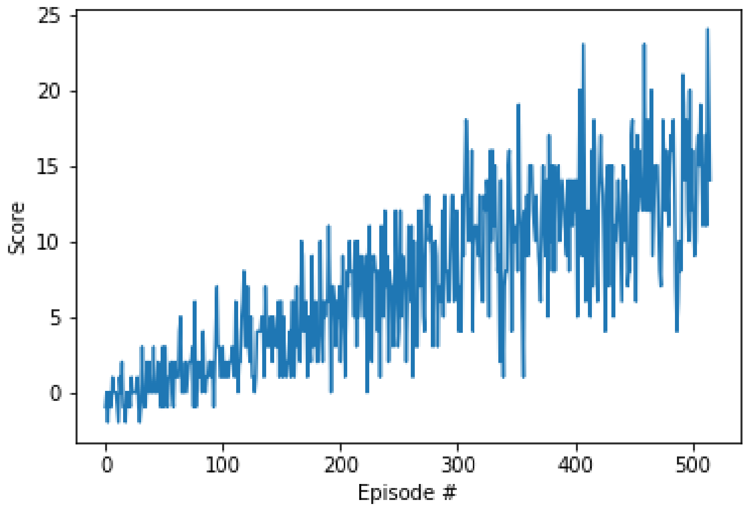
The same model architecture was obtained from DQN project source code in Udacity workspace.

- The neural network consists of three fully connected layers: two hidden layers and one output layer.
- The first two connected layers are with ReLU activation function.
  a. The first layer maps the input state of 37 states with hidden layer of 64 hidden units.
  b. The first hidden layer of 64 hidden units is fully connected with the second hidden layer of 64 hidden units.
- The last layer is a fully connected layer with Linear classification.
  a. The second hidden layer is connected to the output layer of 4 actions.

Plot of Rewards

The plot shows the environment solved in 416 episodes with average score of 13.04.

```
Episode 100      Average Score: 0.75
Episode 200      Average Score: 3.92
Episode 300      Average Score: 7.34
Episode 400      Average Score: 10.74
Episode 500      Average Score: 12.53
Episode 516      Average Score: 13.04
Environment solved in 416 episodes!      Average Score: 13.04
```



Ideas for Future work

To check the possibility of approaches to converge with average score greater than 13 in an early episode:

- Change the hyperparameters as for the batch size, the learning rate, and the discount factor.
- Change the network architecture in terms of the number of hidden layers and hidden units per each hidden layer.
- Use different strategies that can provide improvements to Deep Q Learning **[1]** and with referring to Udacity course such as:
  - Double DQN
  - Prioritized Experience Replay
  - Dueling DQN

References

1. https://medium.freecodecamp.org/improvements-in-deep-q-learning-dueling-double-dqn-prioritized-experience-replay-and-fixed-58b130cc5682