Navigation Project Report

1. Learning Algorithm

(1) The experience replay

Keeping track of a replay buffer and using experience replay to sample from the buffer at random, we can prevent action values from oscillating or diverging catastrophically. The replay buffer contains a collection of experience tuples. The tuples are gradually added to the buffer as we are interacting with the environment. The act of sampling a small batch of tuples from the replay buffer in order to learn is known as experience replay. In addition to breaking harmful correlations, experience replay allows us to learn more from individual tuples multiple times, recall rare occurrences, and in general make better use of our experience.

(2) Deep-Q network

For most problems, it is impractical to represent the -function as a table containing values for each combination of and . Instead, we train a function approximator, such as a neural network with parameters , to estimate the Q-values. Because the neural network will approximate the Q-values, it can also generalize to unexplored states. This type of neural network is called a Deep Q Network (DQN).

Consequently, the neural network requires only a “state” as the input to return a Q value for each possible action at the state. This single input is in contrast to a Q-Table — you need to provide both the state and the action to return the Q value.

(3) Hyperparameters

BUFFER_SIZE = int(1e5) # replay buffer size

BATCH_SIZE = 64   # minibatch size

GAMMA = 0.99   # discount factor

TAU = 1e-3   # for soft update of target parameters

LR = 5e-4   # learning rate

UPDATE_EVERY = 4   # how often to update the network

Deep network structure:

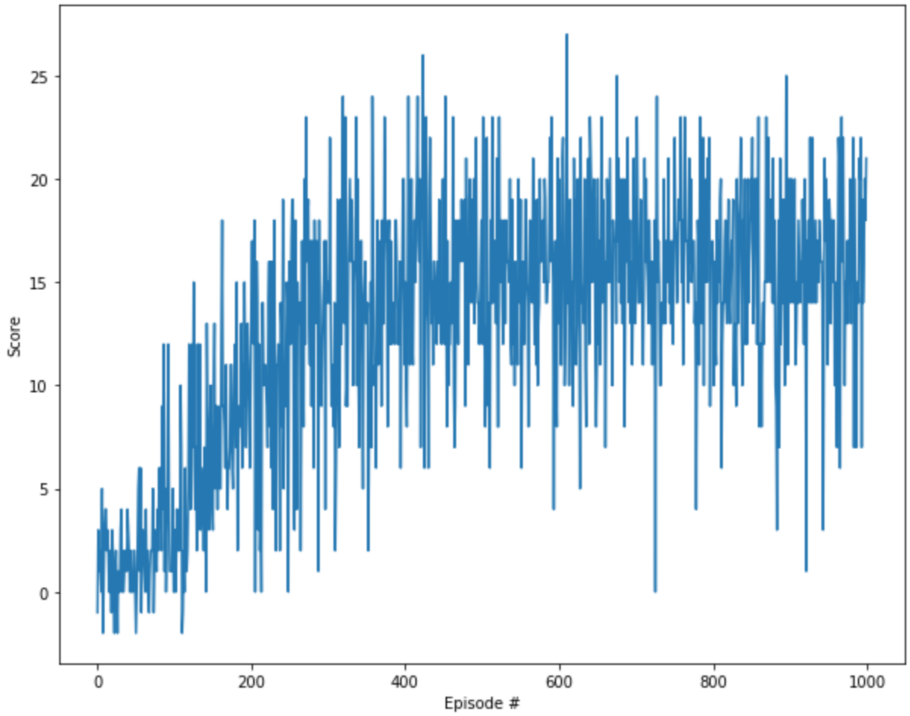
There are two hidden layers.

The hidden layer sizes are 64 and 64 respectively.

The activation function is relu.

2. Plot of Rewards

The plot of rewards shows that after about 1000 steps. The average of rewards is about 15.



The required episodes for reaching the desired score is 400.

3. Ideas for Future Work

(1) Improving the structure of the deep network. For example, by Adding CNN layer.
(2) Tuning the hyperparameters