Project 1 Technical Report: Navigation

Soroosh Rezazadeh

This report is a brief technical description of the Navigation project in for the Deep Reinforcement Learning Nanodegree from Udacity.

Environment

We use an environment built in Unity. The environment brains are responsible for deciding the actions of their associated agents. The characteristics of the environment is shown below.

```
Unity Academy name: Academy
Number of Brains: 1
Number of External Brains : 1
Unity brain name: BananaBrain
Vector Observation space type: continuous
Vector Observation space size (per agent): 37
Number of stacked Vector Observation: 1
Vector Action space type: discrete
Vector Action space size (per agent): 4
```

Learning Algorithm

We use Deep Q-Network (DQN) to solve this problem as explained in the video lectures. In the DQN algorithm the external loop by episodes parameter is executed until the maximum number of episodes, e.g. 2000 here, is reached or the last 100 achieved episode scores in the deque would be greater or equal than 13.
The internal loop gets the current action from the agent, and using this action DQN algorithm acquires state and reward from Unity environment env. The agent has *state*, *action*, *reward*, *next_state*, and *done* parameters for the next training step.

Results:

We trained the agent four times with different hyperparameters and got the best average test score for the following set of parameters:

```
Train: 3, Test: 0, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 15.0
Train: 3, Test: 1, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 16.0
Train: 3, Test: 2, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 11.0
Train: 3, Test: 3, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 22.0
```

```
Train: 3, Test: 4, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 19.0
Train: 3, Test: 5, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 20.0
Train: 3, Test: 6, Episode: 573, fc1_units: 64, fc2_units: 64, eps_start: 0.992, Score: 15.0
Average Score:   16.8571428571
```

The training session for this set of parameters is brought below. The plot below shows the rewards per episode over last 100 episodes.

```
fc1_units:  64 , fc2_units:  64
train_numb:  3 eps_start:  0.992
Episode: 573, elapsed: 0:09:11.075954, Avg.Score: 13.04,  score 23.0, How many
scores >= 13: 59, eps.: 0.10
 terminating at episode : 573 ave reward reached +13 over 100 episodes
```
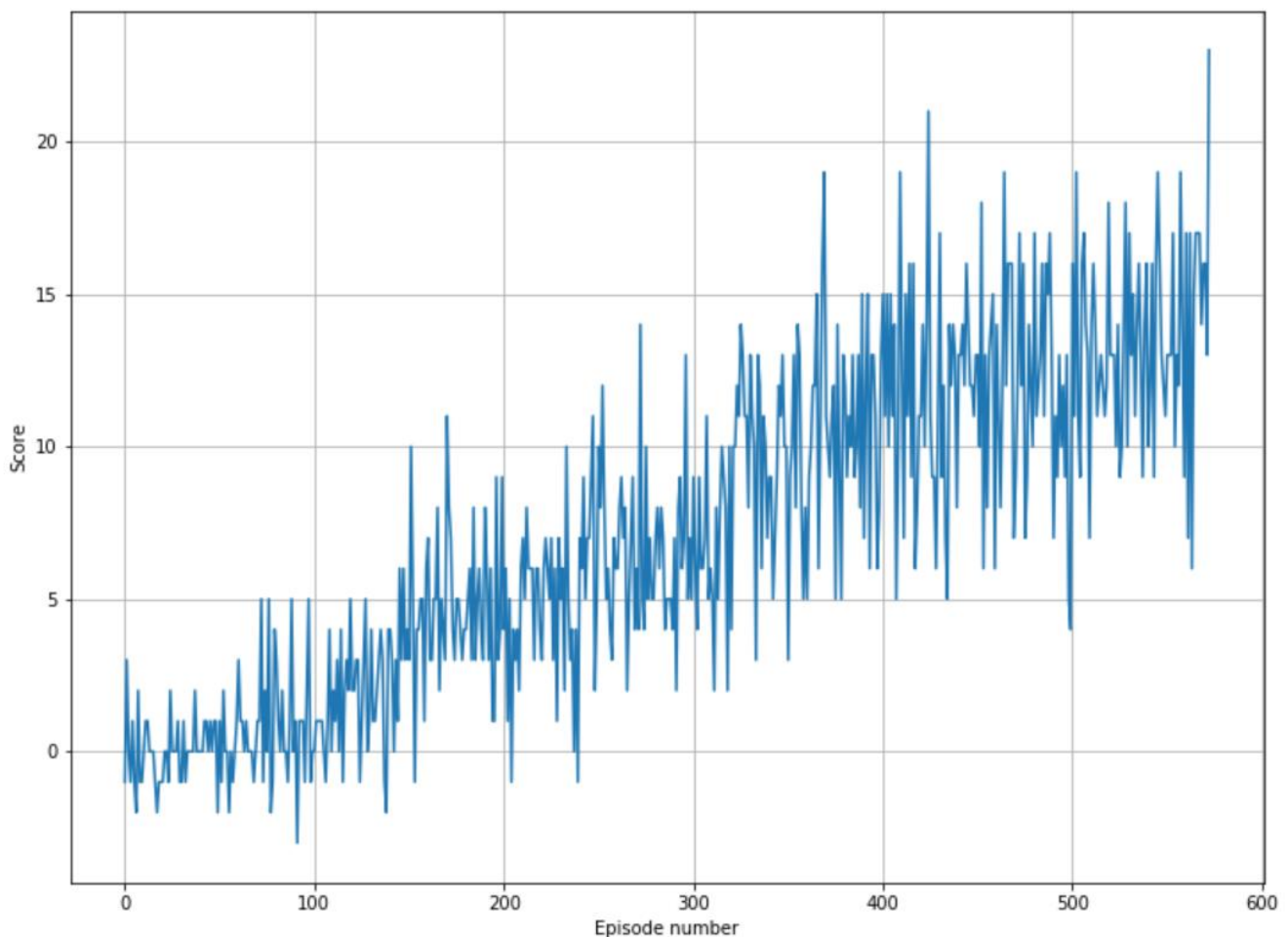

Future Work

To improve the performance:
- We can explore implementing Double DQN, Dueling DQN, and Prioritized Experienced Replay.
- Adding one or more layers to the neural network
- Do more hyperparameter tuning on epsilon and fc1_nodes and fc2_nodes