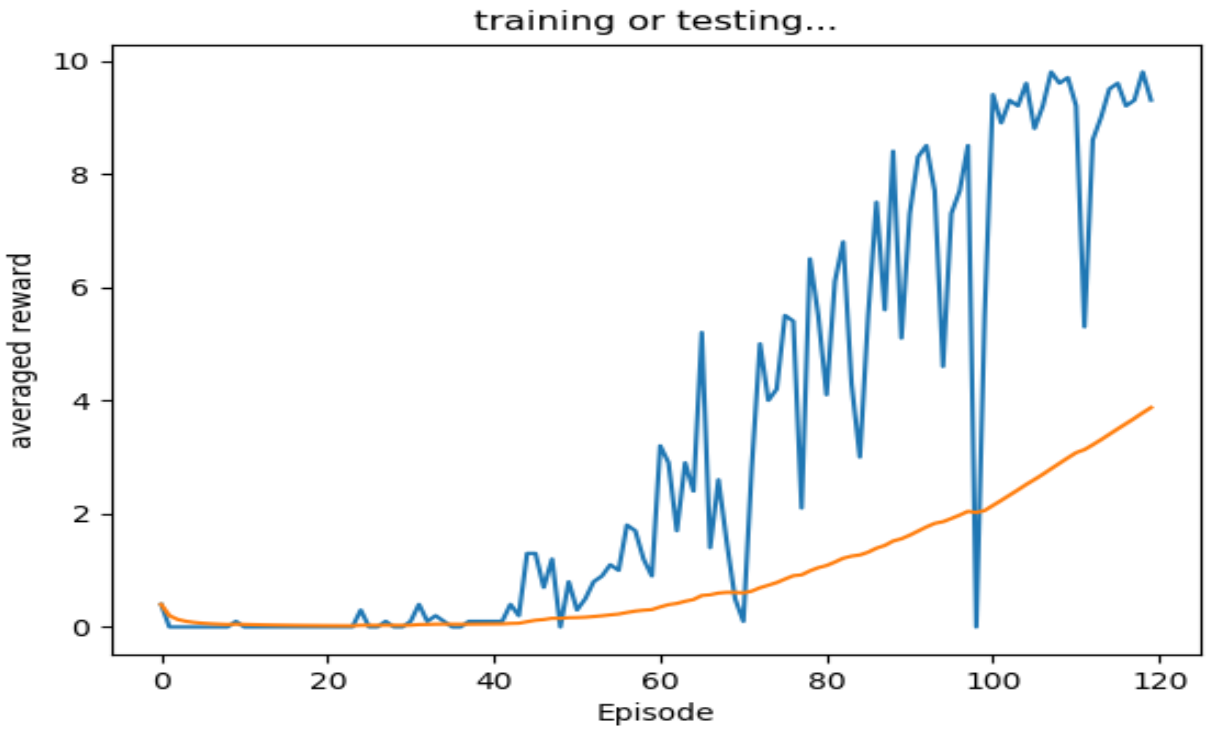
Choo Han Ye Samson 1002439

**AI Coding Homework Report**

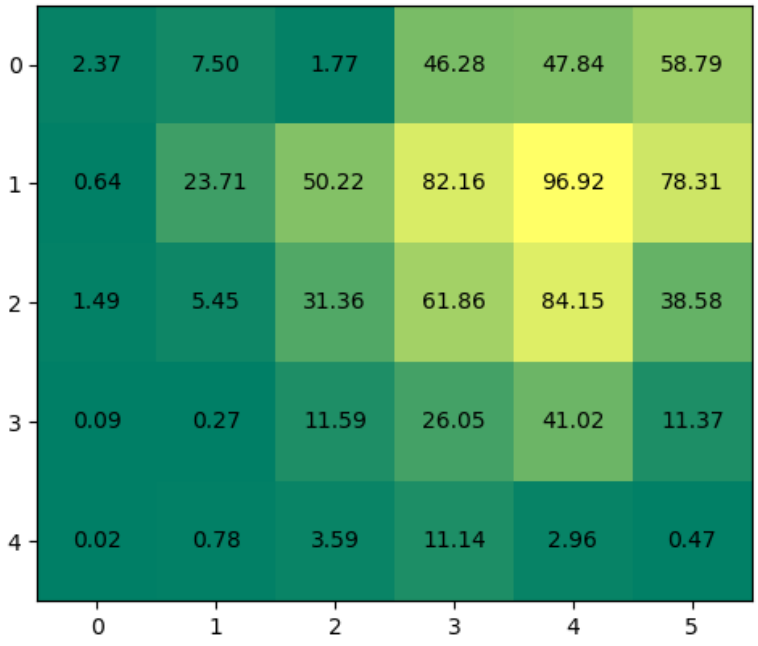
Task 1

1. trainsth()

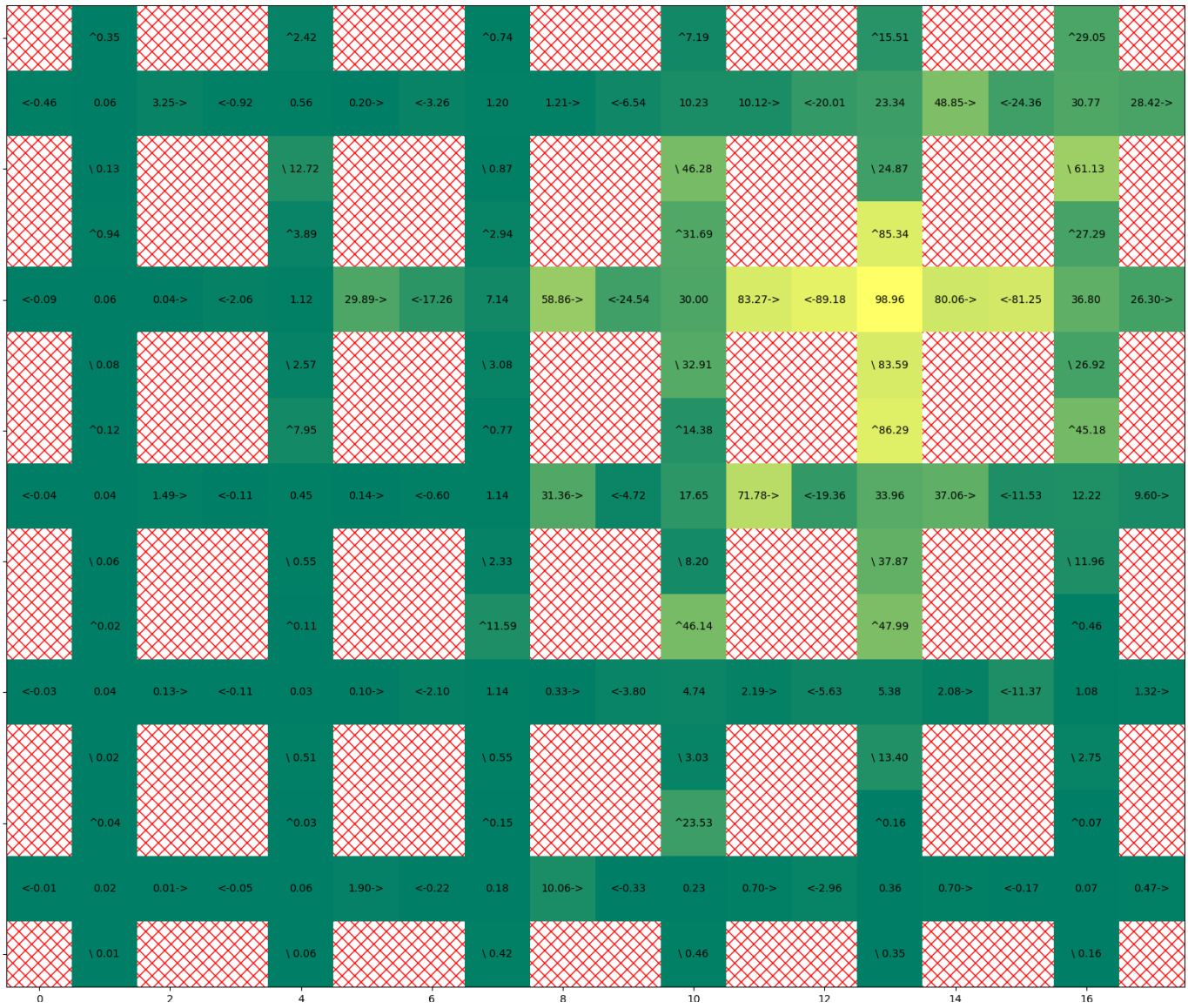
Below is the graph of the averaged reward over the number of episodes:




Below is the graph of the values:



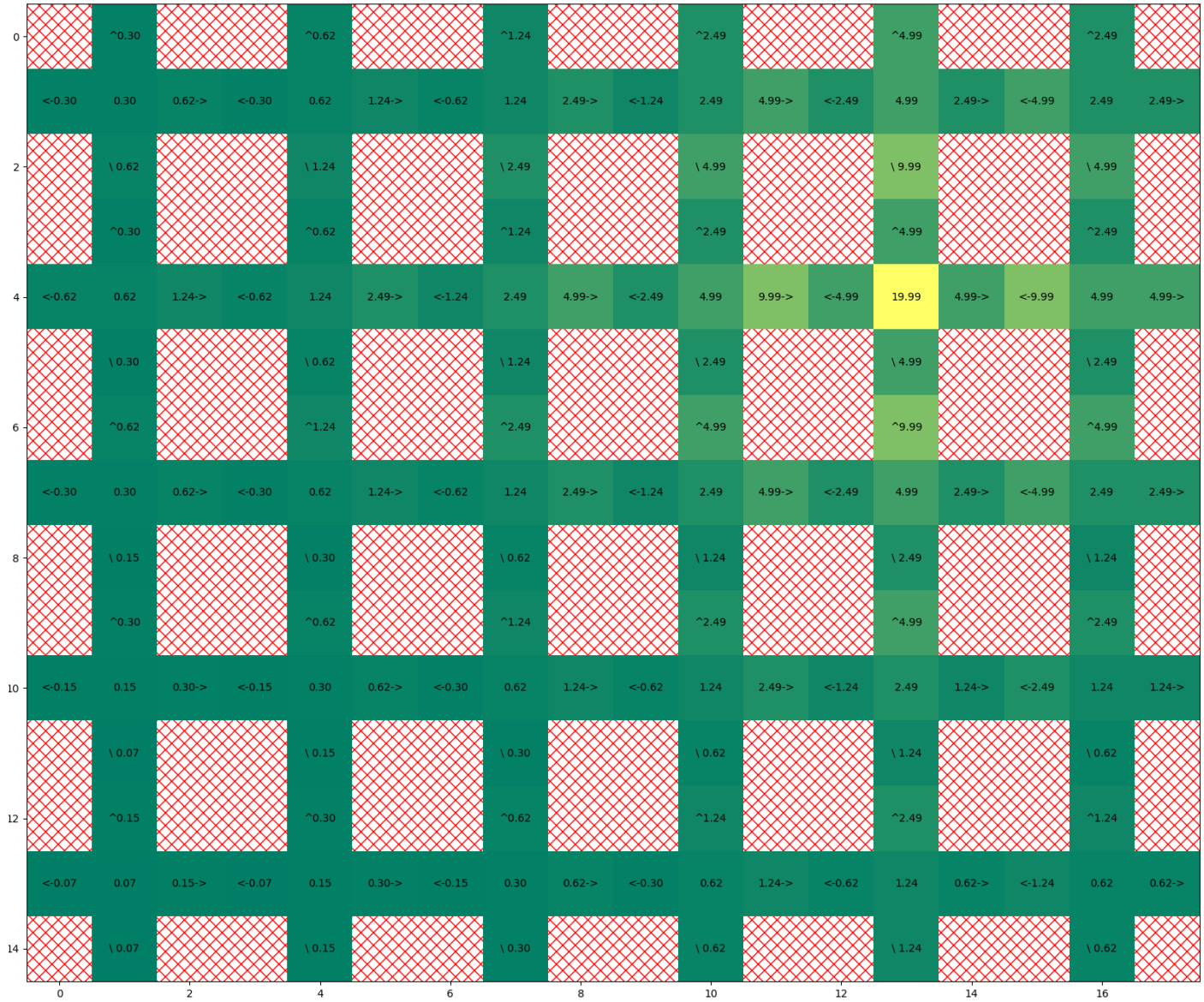
Choo Han Ye Samson 1002439
Below is the graph of Q(s,a):



Choo Han Ye Samson 1002439

2. runmdp():

Below is the graph for Q(s,a):




Task 2

Below are the updated parameters as suggested:

BATCH_SIZE = 128

GAMMA = 0.999

EPS_START = 0.9

EPS_END = 0.02

EPS_DECAY = 200

TARGET_UPDATE = 30

EPS_END_STEPS = 12000

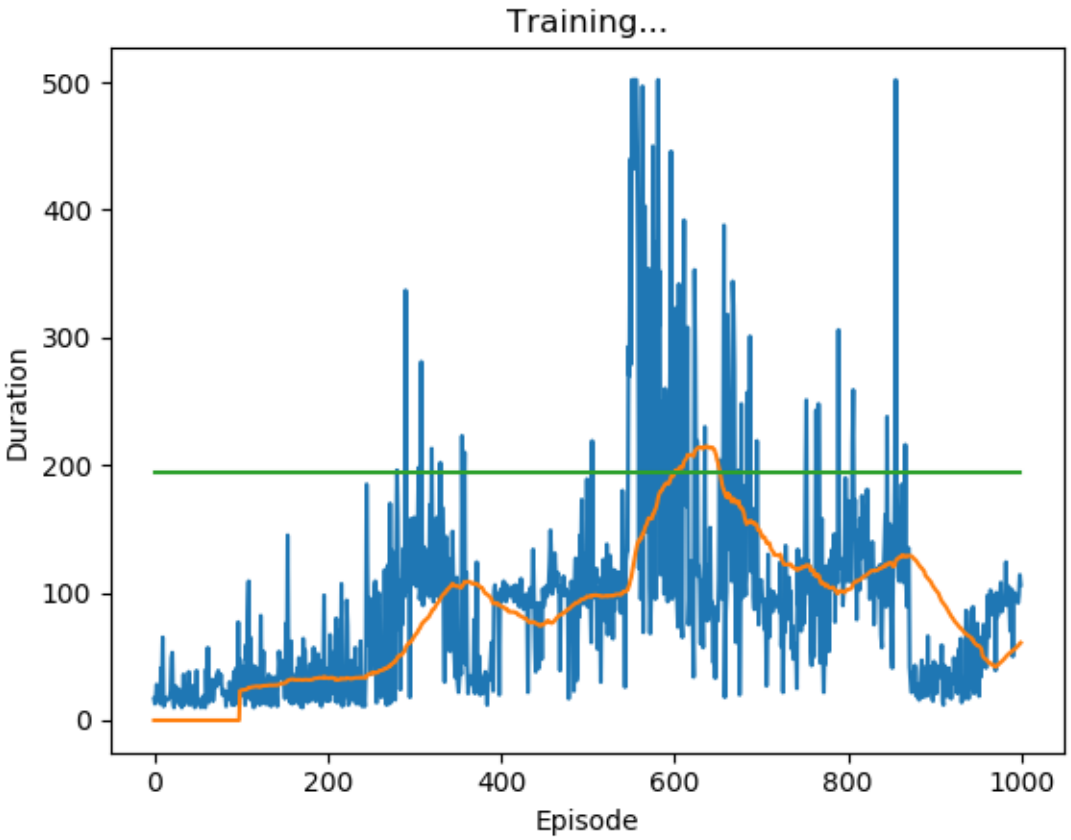
REPLAY_MEMORY_SIZE = 50000

LEARNING_RATE = 0.005

WEIGHT_DECAY = 0.000005 #how much is mild?

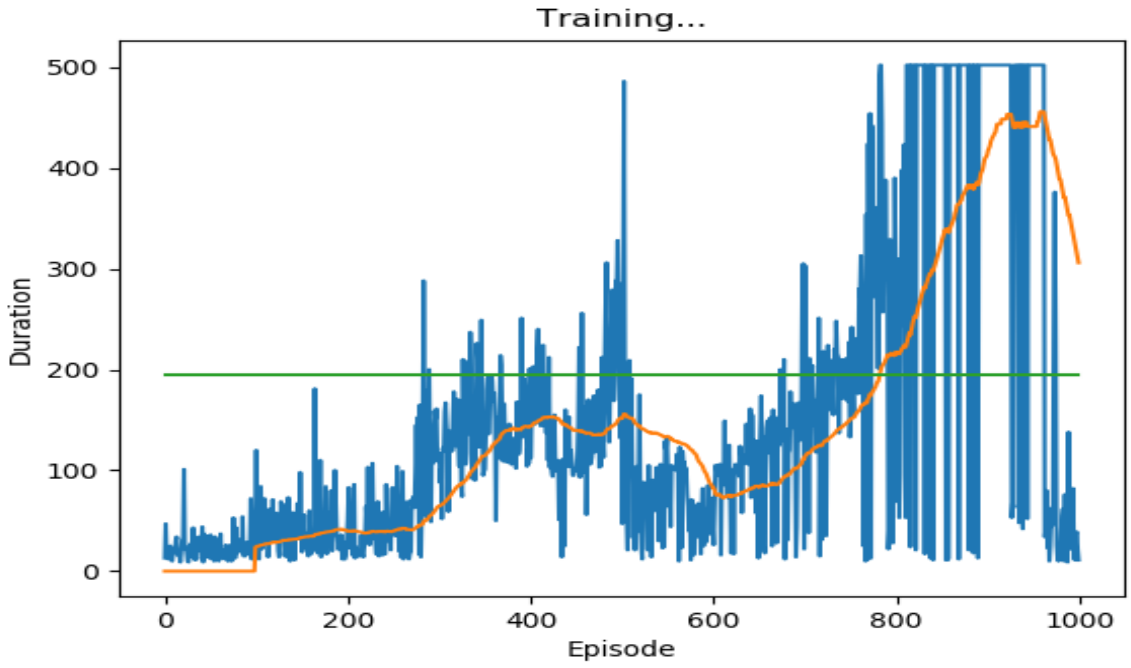
Choo Han Ye Samson 1002439
The number of steps per episode has also been limited to 500, and each of them is ran 5 times.

This is the baseline result:




It reaches the target duration (195) first at around 300$^{th}$ episode, and the averaged duration achieved target at around 600$^{th}$ episode. The maximum averaged duration over the past 100 episodes achieved is 214.6300.

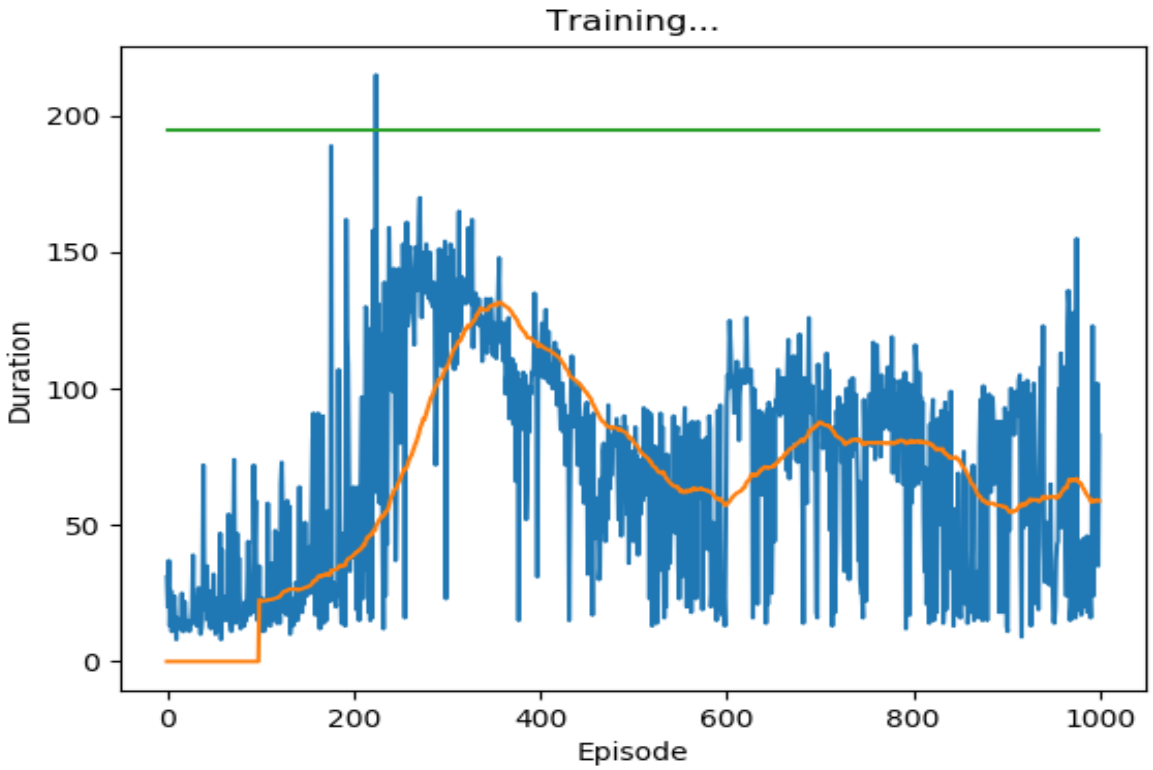
MSE loss instead of huber loss:



Choo Han Ye Samson 1002439

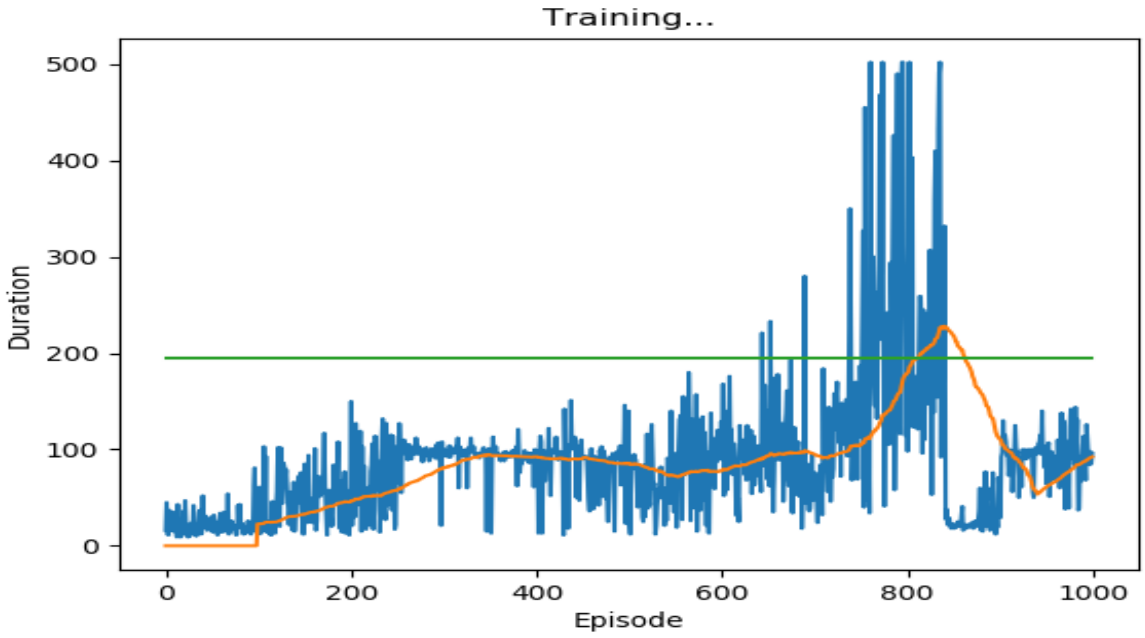
Surprisingly MSE Loss performed much better. It reached 195 at around 800th episode, and the highest averaged reward achieved was 455.69.

Replay Memory Size = 30000 (from 50000):



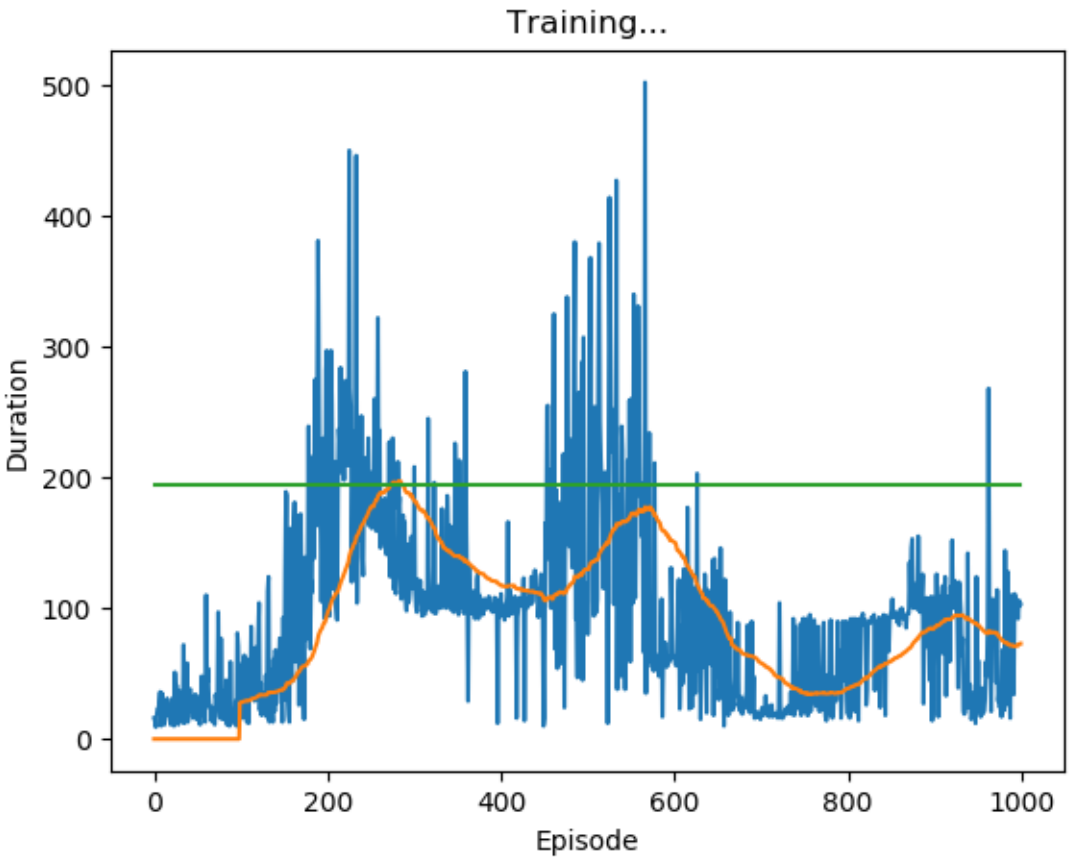

The performance is bad. The highest achieve duration was 131.7600.

Replay Memory Size = 60000 (from 50000):



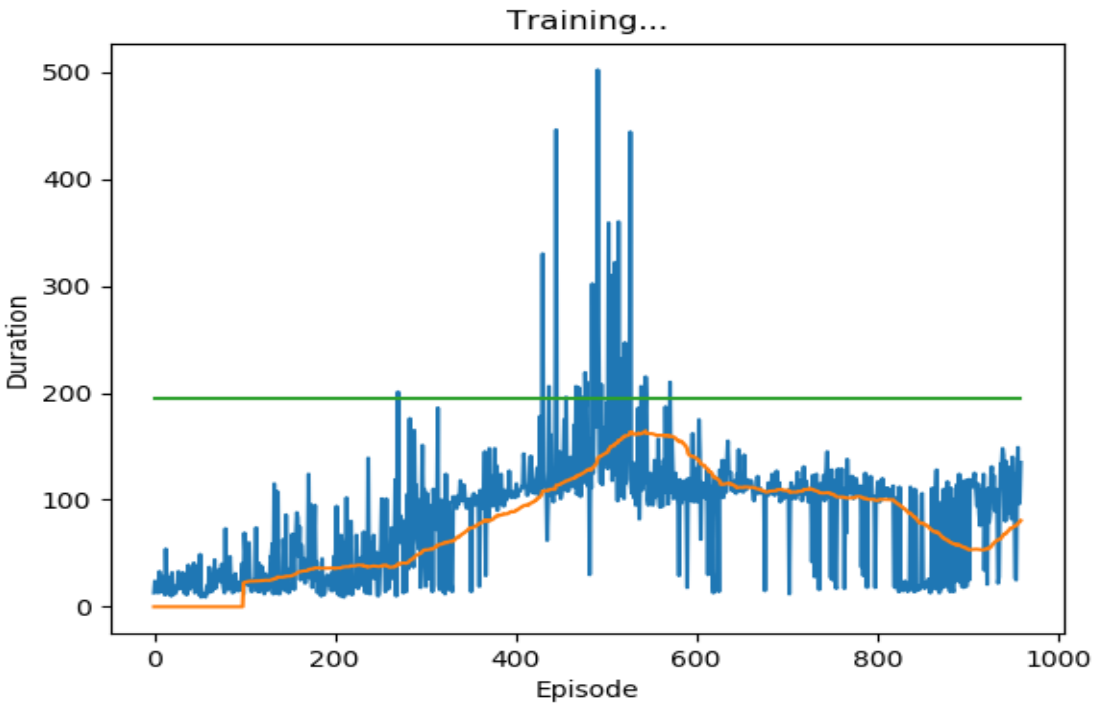

It achieved the target at around 800th episode, and the best performance was 228.0300 average reward. From the 3 data points, we may naively hypothesise that the higher the Replay Memory Size, the better the performance.

Choo Han Ye Samson 1002439
Learning rate = 0.001 (from 0.05):



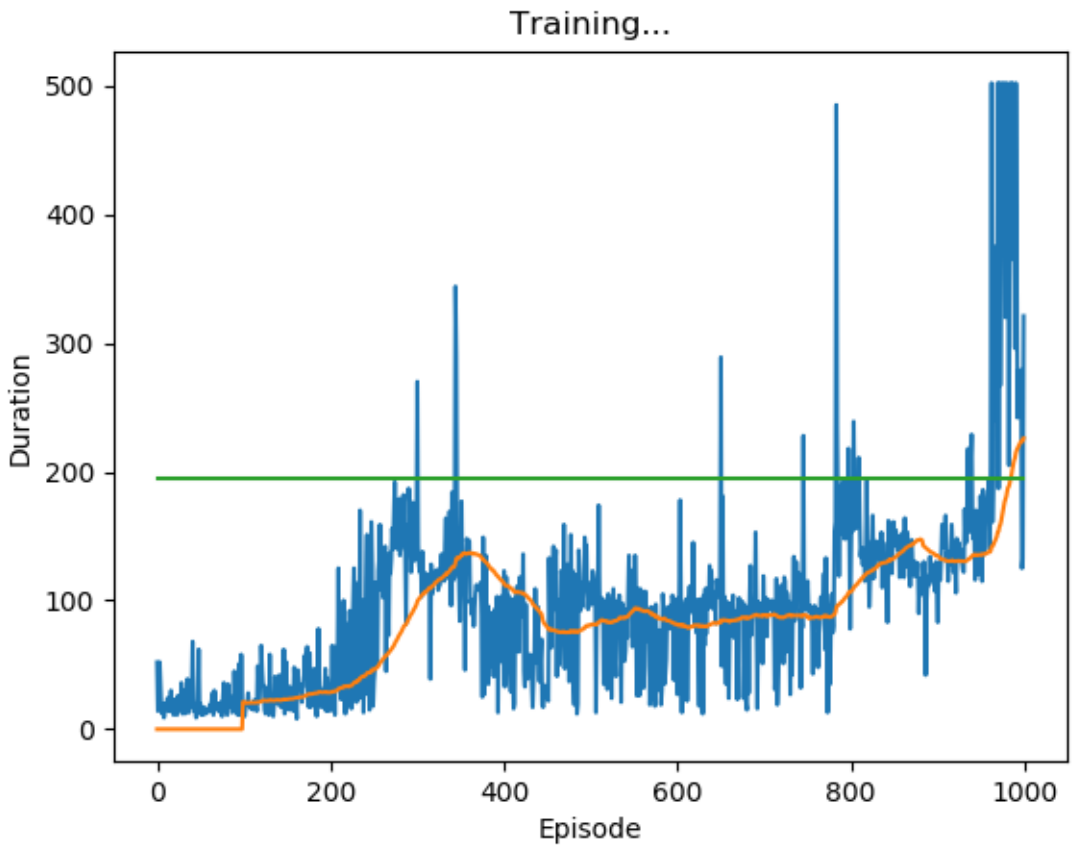

Maximum duration was 197.27, achieved at around 300th episode.The performance was worse, but it achieved higher result early on and managed to reach the target earlier than the other settings.

Learning rate = 0.01 (from 0.05):



Choo Han Ye Samson 1002439
It could not reach the target performance…

Change nn intermediate nodes from 24 to 30:




Having a wider neural net caused it to converge later, towards the end of the 100 episodes. It achieved a maximum performance of 226.0100 at the 100$^{th}$ episode, but it can probably go higher with 50~100 more episodes.

Task 3:

The parameters used are as below:

BATCH_SIZE = 128

GAMMA = 0.999

EPS_START = 0.9

EPS_END = 0.02

EPS_DECAY = 200

EPS_END_STEPS = 20000

TARGET_UPDATE = 30

REPLAY_MEM_SIZE = 50000

LEARNING_RATE = 0.001

Choo Han Ye Samson 1002439

Here is the score of the agent, averaged over the past 100 episodes. It first got a score of 175 at around 300$^{th}$ episode, and then later on the average score reached the target at 1800$^{th}$ episode.