DLRND 1 Project Report

Ariel Kwiatkowski

April 2020

1 Methods and results

The approach I used is relatively simple – it's a DQN algorithm using an MLP model as the Q network. Including the input and output sizes, the neural network layers are: 37, 64, 64, 4. I'm using a separate target network, which is soft-updated to the main network with $\tau = 0.001$. The main network is optimized using the Adam algorithm with the learning rate $5 \cdot 10^{-4}$, performing a gradient update every 4 steps, using a batch of 64 transitions taken from a replay buffer of maximum size of 10000 transitions.

The graph of rewards can be seen in Figure 1, with the heightened standard of requiring a mean reward of 15 (rather than 13). This objective has been achieved after 637 episodes (for a total of 737 episodes).
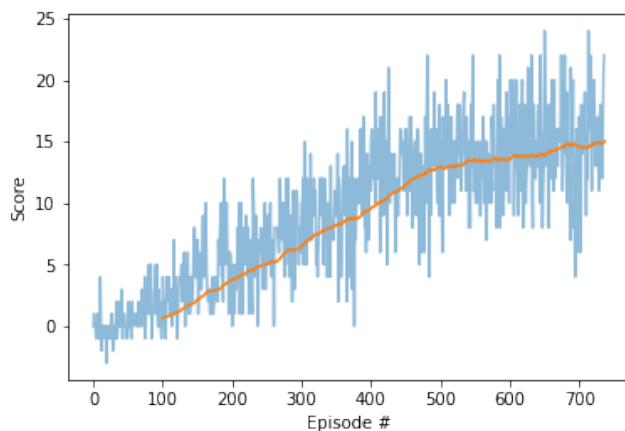

Figure 1: Rewards obtained by the agent, along with the 100-episode running mean

2 Possible improvements

Despite the relatively good performance, it would be possible to increase it further using some of the standard DQN tricks, including using a dueling architecture, or adapting the target network into a double DQN. A more extensive hyperparameter search would probably also be beneficial and allow me to find a network that performs even better.