RL Coursework 1: Question 3

Monday 28th March, 2022

Question 3: Understanding the loss

DQN loss increases as the training progresses primarily due to the non-stationarity (moving target) condition. As the training progresses, the target network gets updated and the average return increases by the sum of rewards, making it harder for the Q-network to approximate to the new average return. In other words, the agent is now selecting better actions, thus the higher values rewards are more frequent, leaving a larger approximation error. It's important to notice that this can be tracked to happen at regular intervals, with lengths are equal to the update frequency: the target gets updated increasing its return values. This explains the spikes we see at these regular intervals as well: the target just moves to a new value which the Q-network wasn't trained to approximate giving out error spikes which decrease as the Q-network starts learning the new target.