Home Work

Erik Norlin

May, 2023

Home work B

1a)

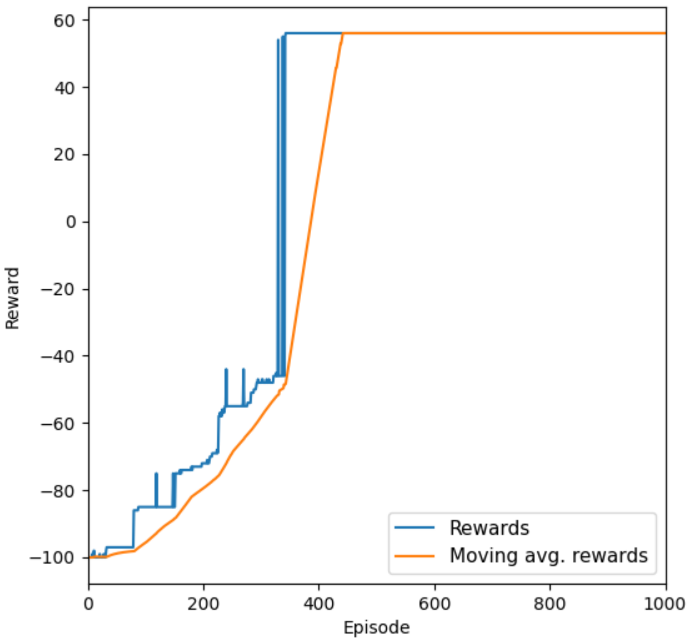

Figure 1: Choosing a random action of all $maxQ$.

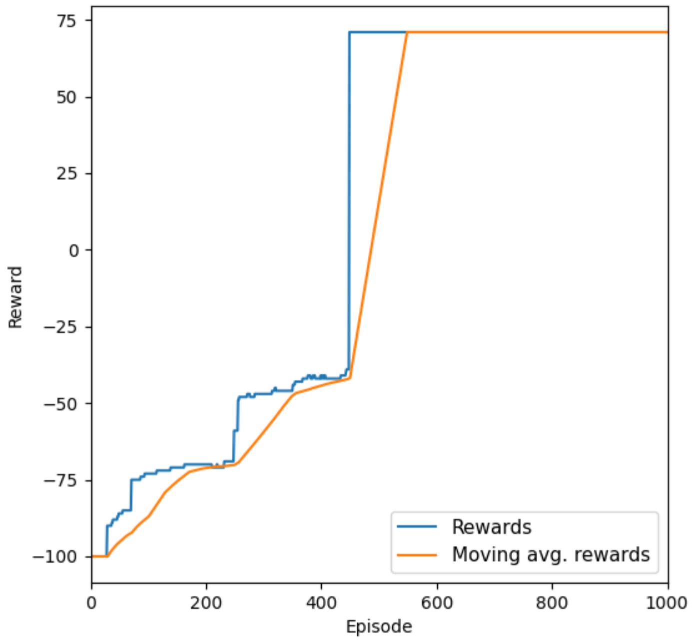

Figure 2: Choosing a deterministic action of all $maxQ$.

1b)

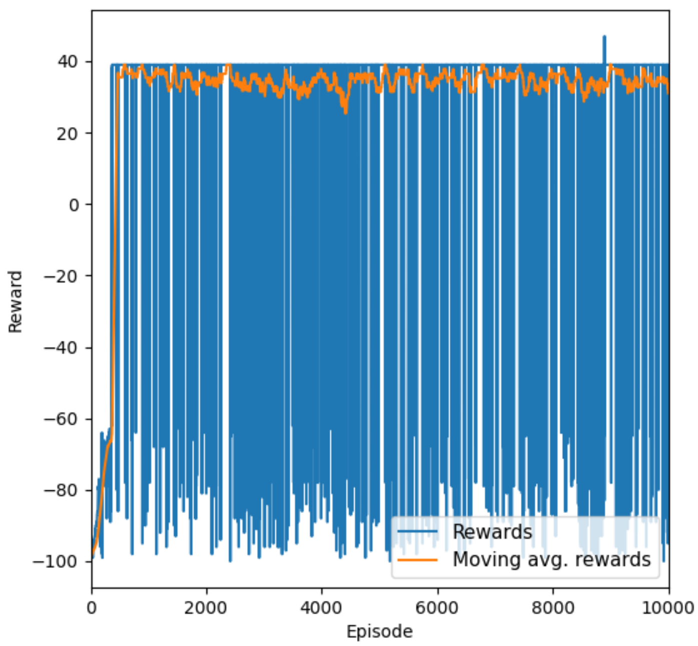

Figure 3: Choosing a random action of all $maxQ$.

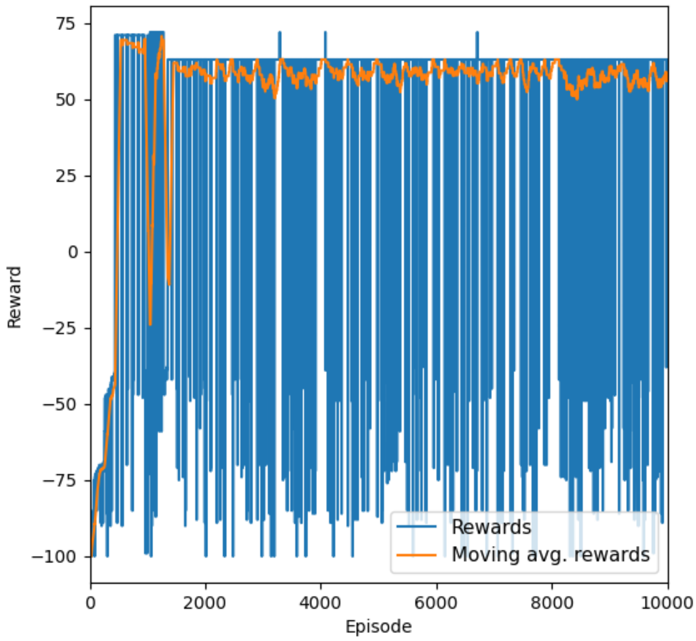

Figure 4: Choosing a deterministic action of all $maxQ$.

Rewards and learning rates are about the same. They perform very similar even though there are about twice as many explored states in b). In b) we have exploration which causes the agent to choose random actions to explore new states, this however causes frequent dips in the rewards that we do not see in a).

1c)

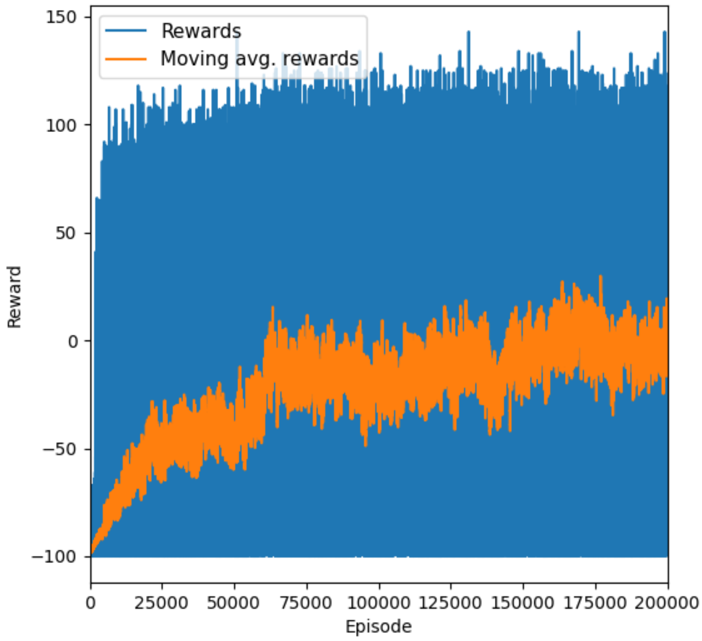

Figure 5: Choosing a random action of all $maxQ$.

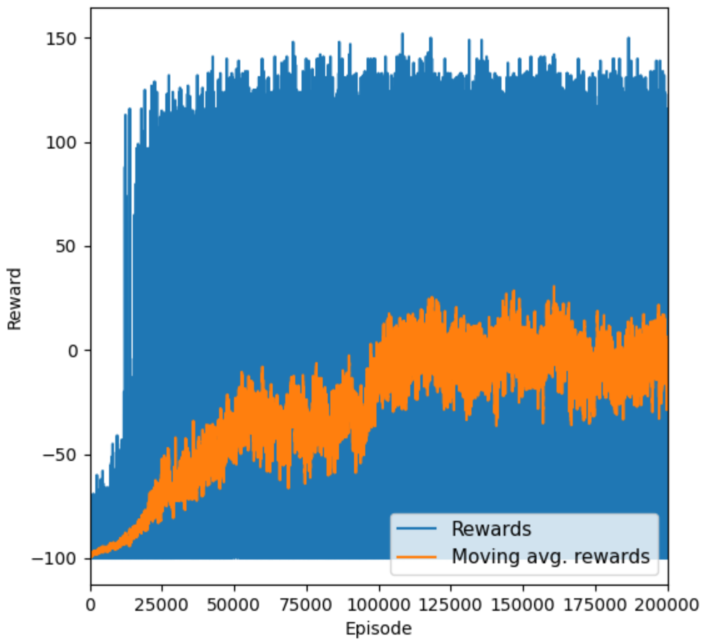

Figure 6: Choosing a deterministic action of all $maxQ$.

The average reward in c) is around 0, much lower than for a) and b). The peaks however are much higher, and there are about 24000 more explored states. The reason for a lower average reward in c) is because the order of the tiles are stochastic where as the order of tiles is deterministic in a) and b). It is easier to train an algorithm with a deterministic pattern, therefore the higher average rewards in a) and b).

1d)

The possibility to implement Q-learning as the game board doubles in size becomes more problematic because of the increasing number of possible states to store in memory. Theoretically, a game board of size 4x4 with 4 different tile types where each square can either be filled or empty can give $2^{4 \cdot 4} \cdot 4 = 262144$ maximal number of states. However, a game board with size 8x8 can give $2^{8 \cdot 8} \cdot 4 = 262144 \approx 7 \cdot 10^{19}$ maximal number of states which is a lot to store in memory. It becomes therefore more suitable to turn to deep Q-learning instead.

2a)

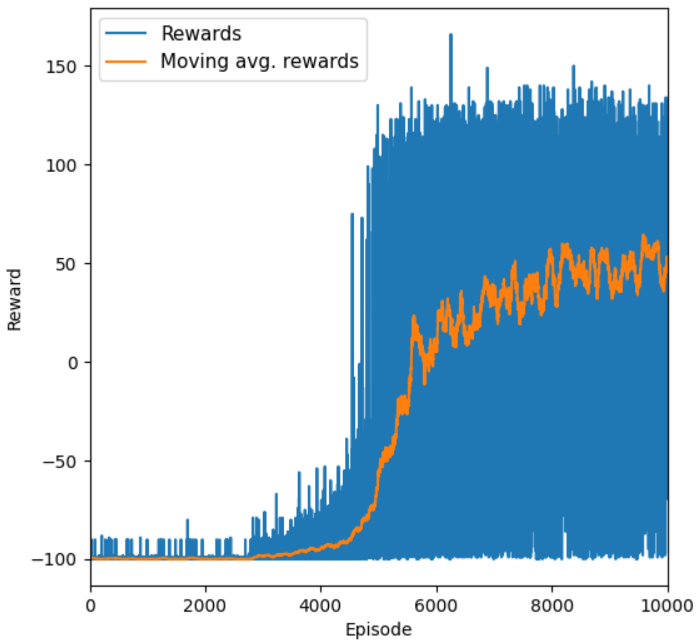

Figure 7: Deep Q-learning.