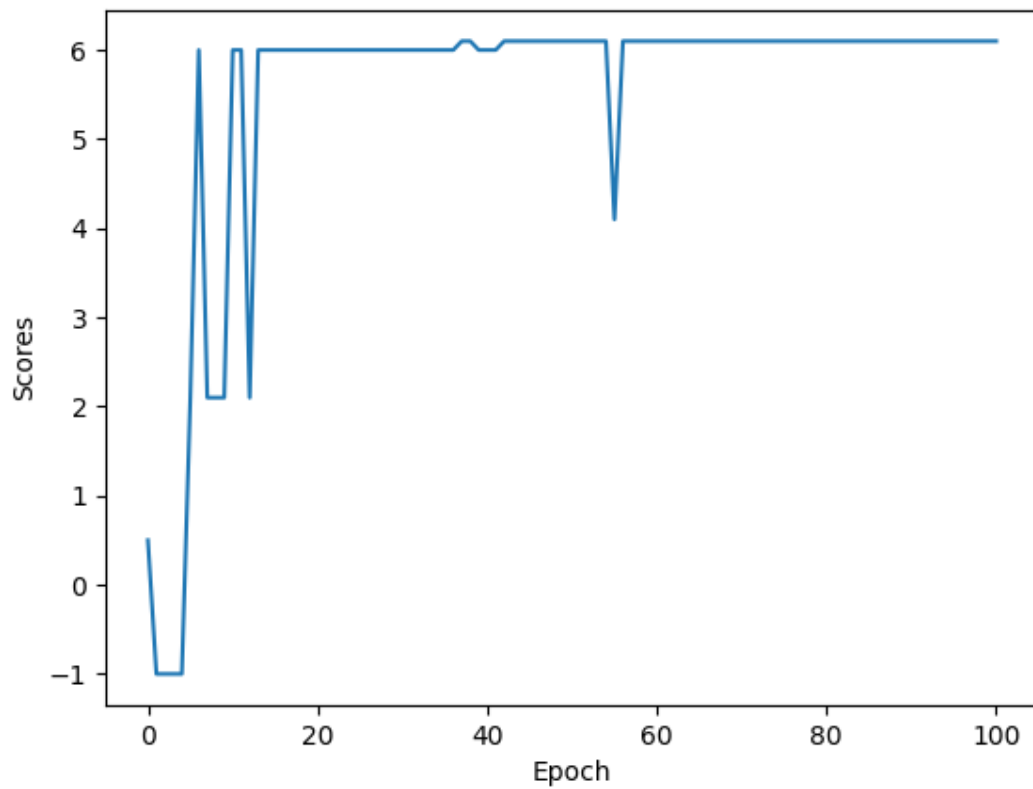


RL homework 2

Writing tasks with attachments

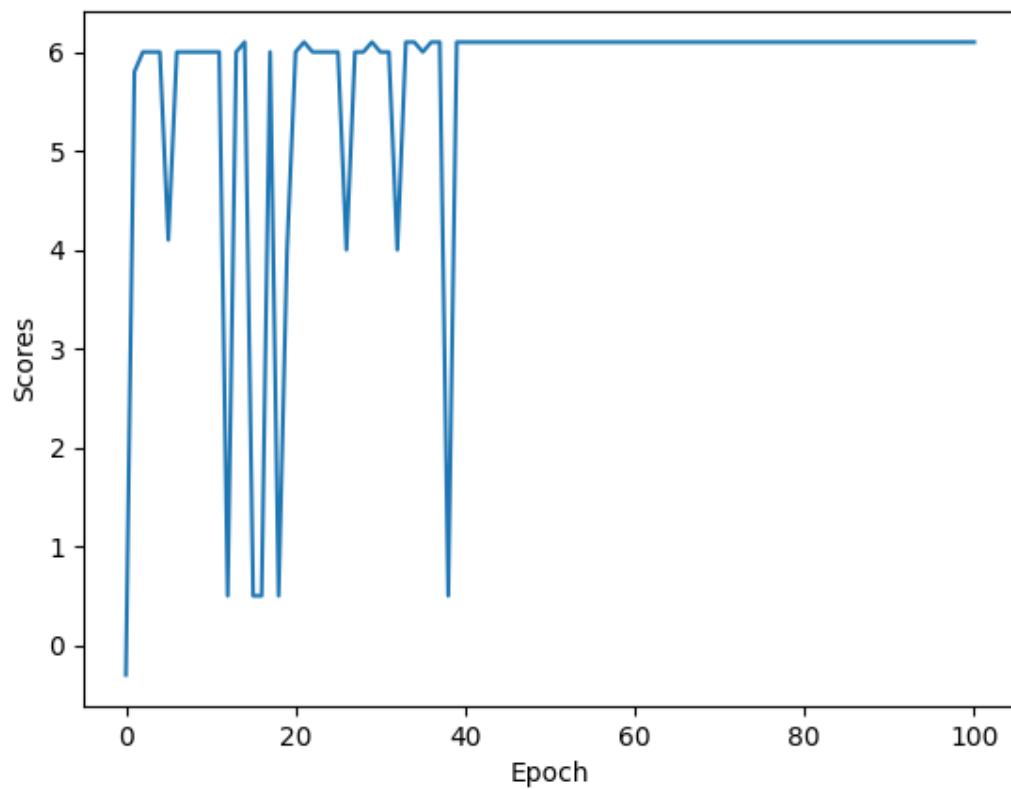
4. Linear Approximation

According to the log, we achieve optimal reward (6.1, as calculated), and scores are as follows:



5. Implementing DeepMind's DQN

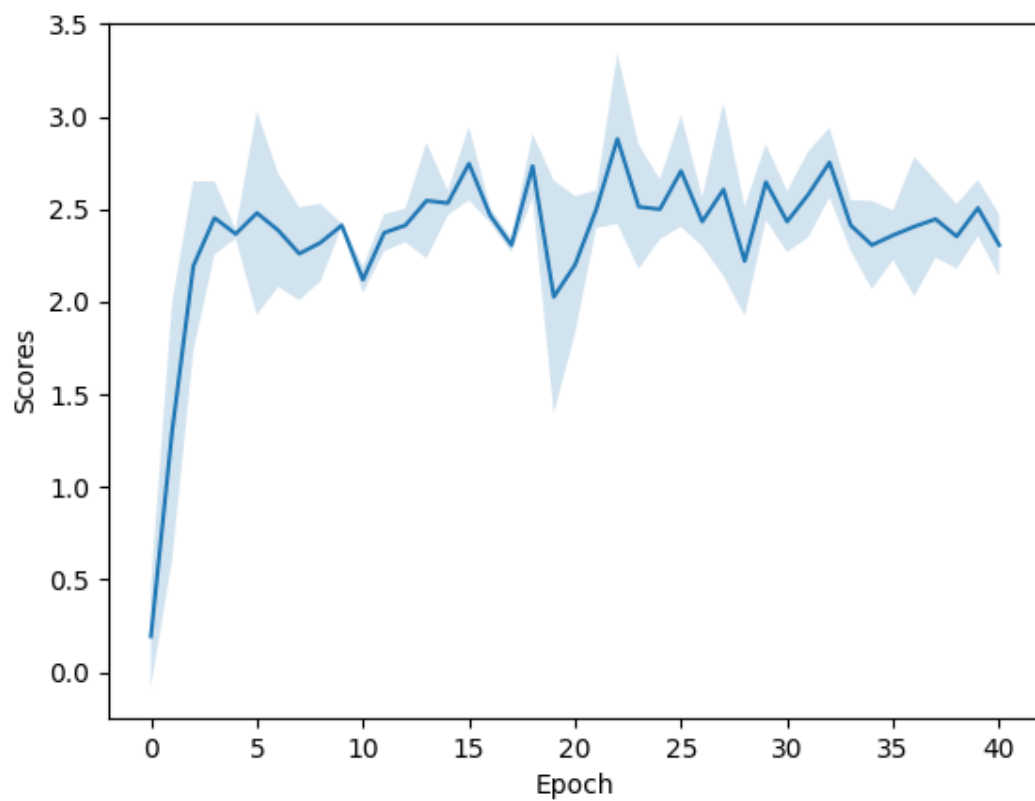
This model shows more amplitude in scores and is more varied. In fact, it reaches optimal reward score earlier, but then draws back. For linear model training time is 1. Minute 38 s, while for DQN is 1 minute 20 s.



6. QDN on MinAtar

a) Linear model

Performance is taking trending course almost immediately – at the 5th epoch we already get result similar to average. Performance improved during the second run, but overall fluctuated around the same value. I don't see if performance could be improved by increasing number of steps, even by heuristic reasons. This is my intuition.



b) DQN

Maximum average reward (on run 2) is 18.64. Maximum average reward is 16.20. Comparing to linear Q value approximator, this performance is more consistent, and it achieves significant bigger reward. It reaches maximum reward and then stays relatively stable. The gap in performance can be explained by adaptiveness of CNN model, while Linear model is bounded by linear configuration.

