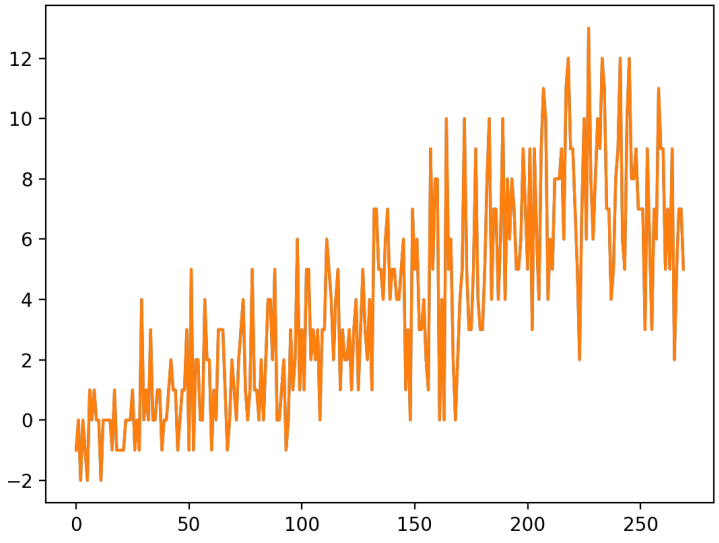

Figure 3: Learning scores per epoch

5 Conclusion and further

We developed a pipeline in order to solve ‘banana‘ environment with Deep Reinforcement Learning. In particular, we implemented **Deep Q-Learning** following the paper specification. Deep Q-learning is an off-policy algorithm which has proven stability and optimal results in multiple tasks. We got the algorithm to hack the environment in less then 300 epochs. As discussed above, we believe the noisy of the learning could be alleviated with further developments of both learning phase and the buffering phase. We would consider integrating **Prioritized Replay** as a way to optimally sample mini-batches from the queue. Further improvements are available in implementing **Double DQNs** strategy in order to disentangle the calculation of Q-targets for best action estimation from Q-values.