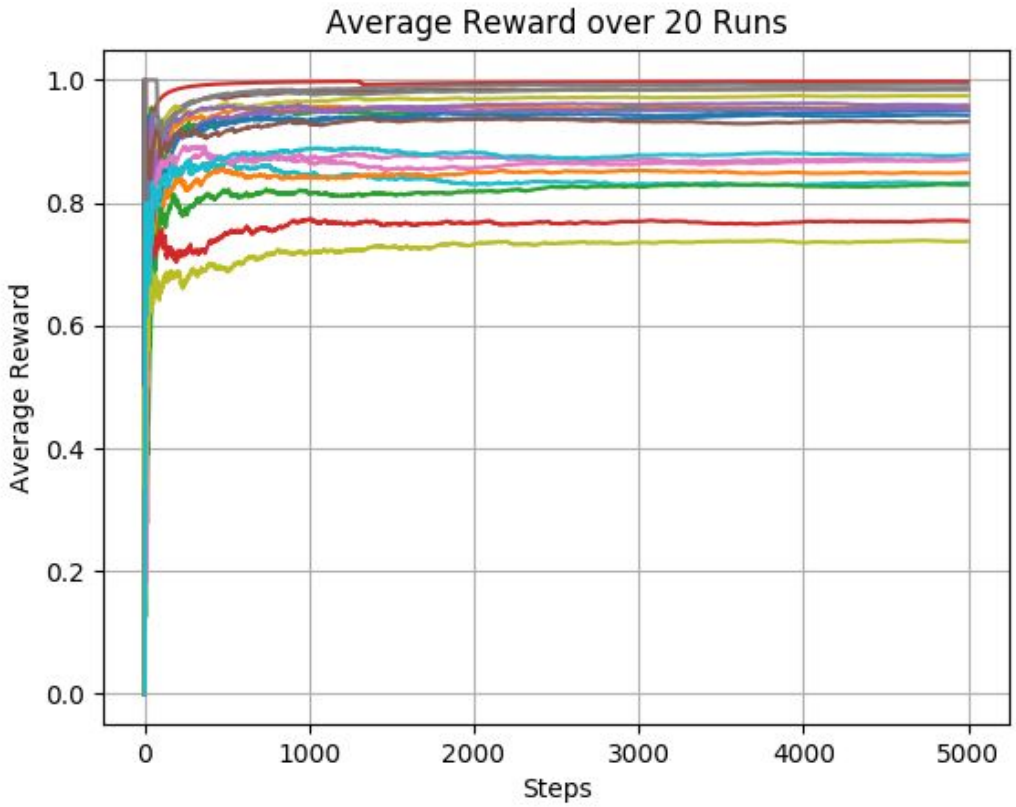
Name:
Nikolas Maier
Oluwatomilayo Adegbite

Student #:
500461990
500569283

Assignment 1: Report 1

The UCB algorithm was great at finding a optimal selection and exploiting it in most runs. Over time with a exploration(C) of 0.01 the algorithm will find the optimal arm or an arm close to the optimal that will bring the average reward above 0.80, for example a run we tested with 100 arms and C=0.01 with 5000 rounds found and exploited an arm with a 96% success rate while missing the optimal arm at a 99% rate but obtained an average reward of 0.94.

If the algorithm has not tried an arm it will test it which usually causes a dip in the average reward while it goes into an exploration phase which can be seen in the graph.



Graph of Average reward of 20 different runs of a 10 armed bandit using the UCB algorithm C=0.01

Because the algorithm uses a record of average previous reward for that arm if a good arm or the optimal arm came up with no reward because of an unlucky pull especially with a low exploration rate the algorithm can miss the optimal arm and settle for a good one that came up with a reward. With a higher exploration rate, the average reward graphs show many more and longer exploration dips while the algorithms looks for arms with greater reward before going back to exploiting. In general, runs where the average reward was low the distribution in the probability of success for the arms was also low as can be seen by the outlier result where the average reward is ~0.64.