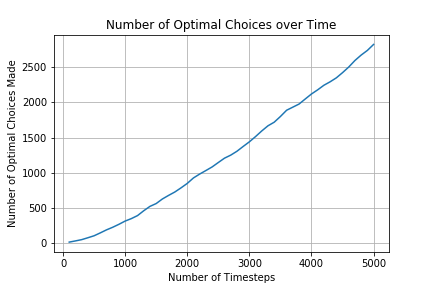
Name: Shaun Mathew

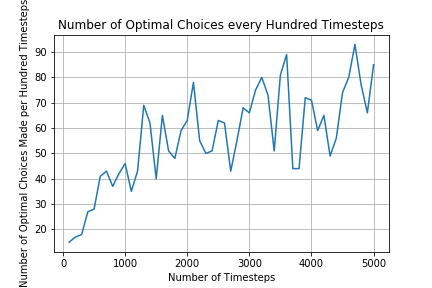
Assignment #1

**Question 1:**

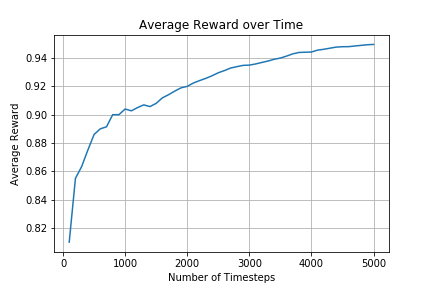
We chose c to be 1 in all runs of our UCB algorithm.

****

As time progresses, the UCB algorithm makes more optimal choices over time. This trend is better seen when we look at the number of optimal choices made every 100 iterations.As we can see below. A large fraction of the actions taken at later timesteps is the optimal one. Therefore, the UCB performs quite well on this task.

****

We now look at the average reward over time. As we can see below, the average reward increases over time. This is to be expected since we see that a large number of choices at later timesteps are the optimal ones.

****

Look at the results from multiple runs in q1\_out\_100\_run.txt, we see that the algorithm tends to achieve a high final reward and picks the optimal action quite often. In cases where the algorithm does not pick an optimal action, we see that the probabilities between the two actions that have high pick counts are similar.