Topics in Machine Learning: Assignment 1

Sahil Chelaramani
20162051

Q1.

We make an assumption that if an arm has a loss = 1, the reward received is 0, and if the loss = 0, the reward = 1.

Hence, we simplify the pseudo-regret to:-

R = T – sum(#rewards in T iterations)

The plot of the pseudo-regret to the total number of steps, as well as, the plot of the number of times each arm is pulled is given below.

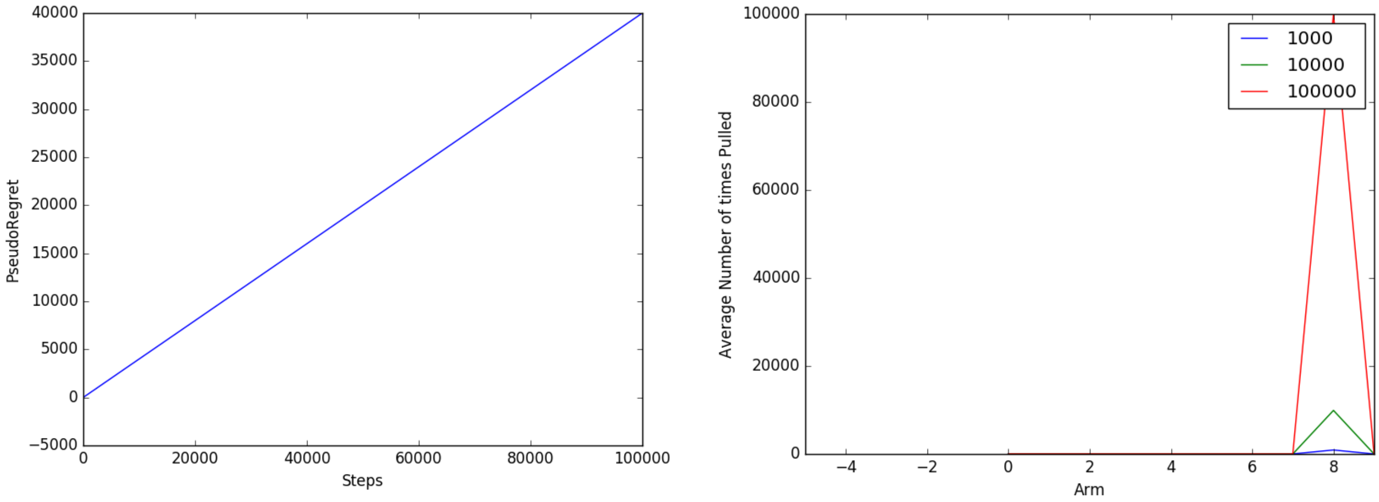

Figure 1*: On the left is a plot of pseudo-regret to number of steps. To the right is the number of times an arm is pulled.*

As can clearly be seen, the 9th arm is pulled frequently. Intuitively, This arm is the best arm, which maximizes the reward since the loss associated with the arm is the least.

Q2.

The solutions to the second question are plotted below. The experiment was run for T = 100000, and averaged over 200 runs.

The values of Episilon = [0, 0.01, 0.1].

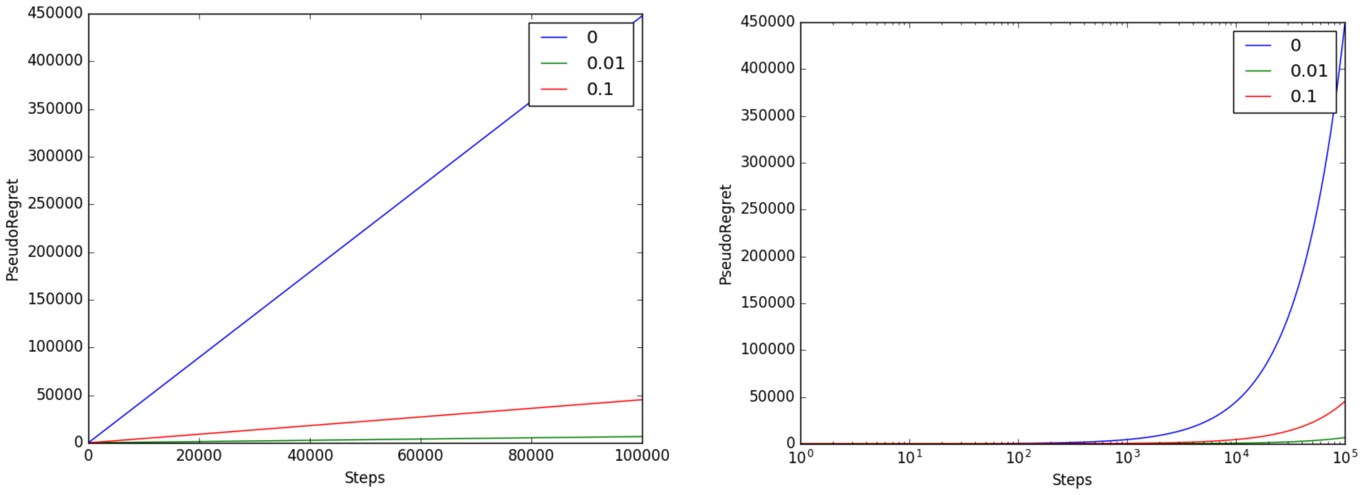

***Figure 2**: On the left is a plot of pseudo-regret to number of steps. To the right is the same plot, with a logarithmic x-scale.*

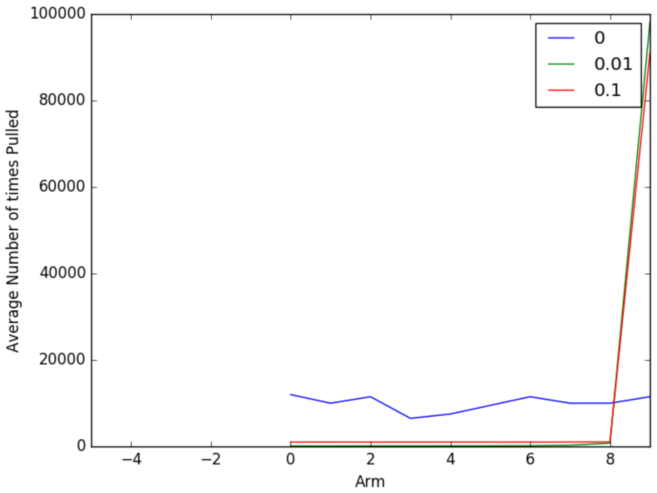

***Figure 3**: Plot of the number of times an arm is pulled.*

As can be easily seen from the above plots, the 10th arm is the optimal arm to be pulled. Intuitively, this is true because it has the maximum average reward(Value = 15) compared to the other arms.

Additionally, notice that the best lowest pseudo-regret is obtained for the value of Epsilon = 0.01. This is probably because the Algorithm with Epsilon = 0, does not explore enough and hence fails to find the best arm due to an initially poor sampling strategy. The case with Epsilon = 0.1 probably spends too much time exploring, and hence results in a regret value which is higher.

Another interesting to notice is, the algorithm with Epsilon = 0, essentially sticks to the same arm without exploring the other arms. Hence, the pseudo-regret generated for this arm is always increasing, and over 200 runs, the average number of times the arms are pulled is constant.

Q3.
The solutions to the third question are plotted below. The experiment was run for T = 100000, and averaged over 200 runs.

The value of C = [0, 0.01, 1, 2].

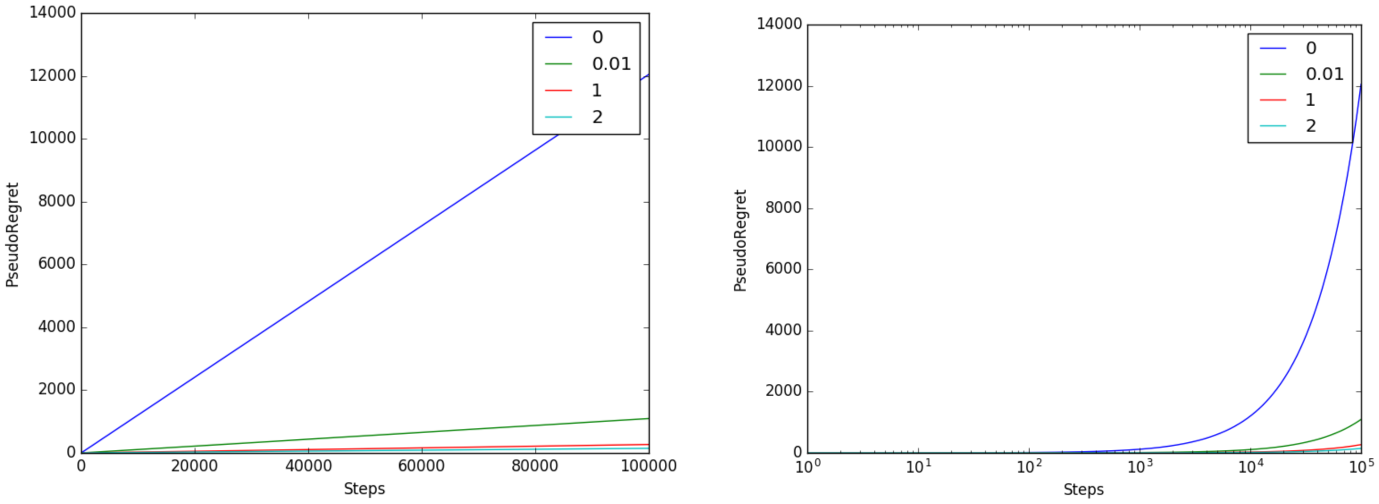

Figure 4*: On the left is a plot of pseudo-regret to number of steps. To the right is same plot, with a logarithmic x-scale.*

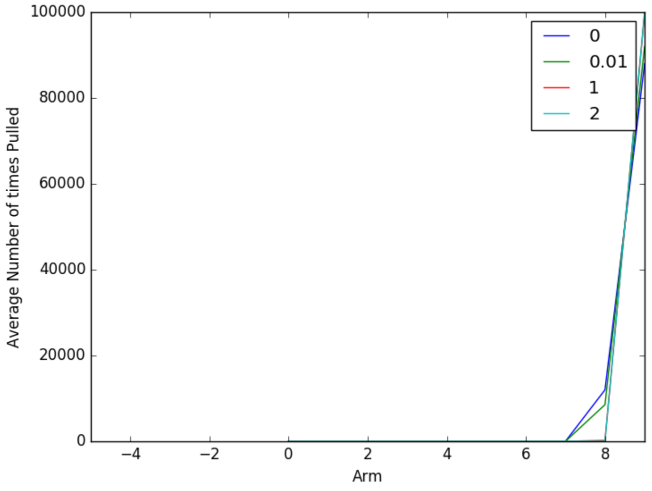

Figure 5*: Plot of the number of times an arm is pulled.*

Similar to the previous question, we find that the algorithm converges to the optimal arm. The UCB algorithm tends to converge much faster to the optimal arm, as can be seen by the substantially lower pseudo-regret levels from the figures.