CS747 - FOUNDATIONS OF INTELLIGENT AND LEARNING AGENTS

ASSIGNMENT REPORT

HARSHIL NAIK (190010030)

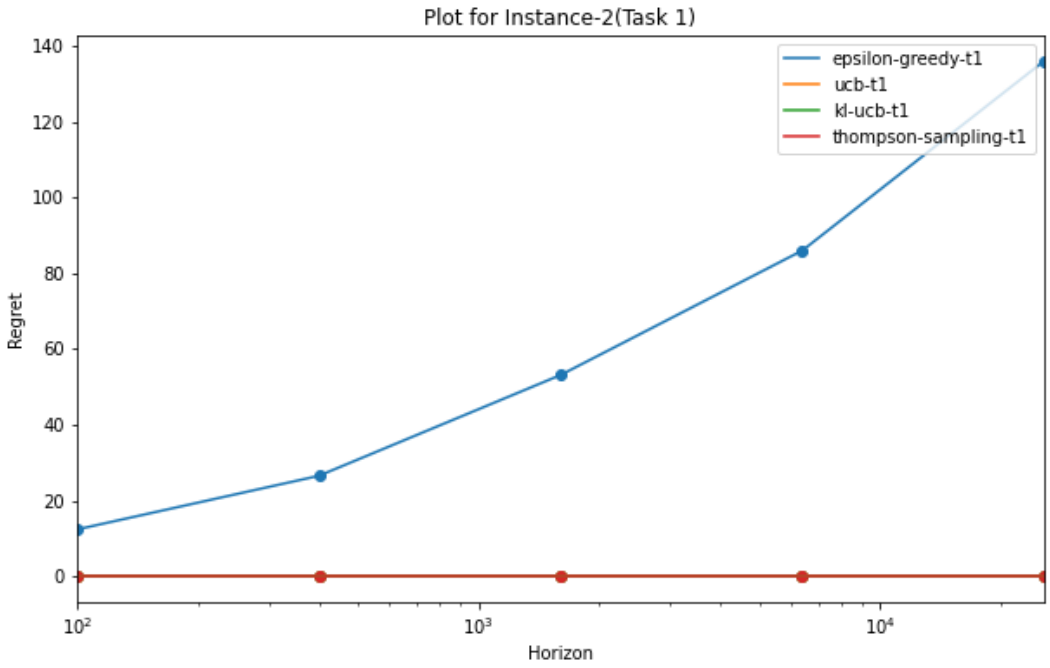
Introduction

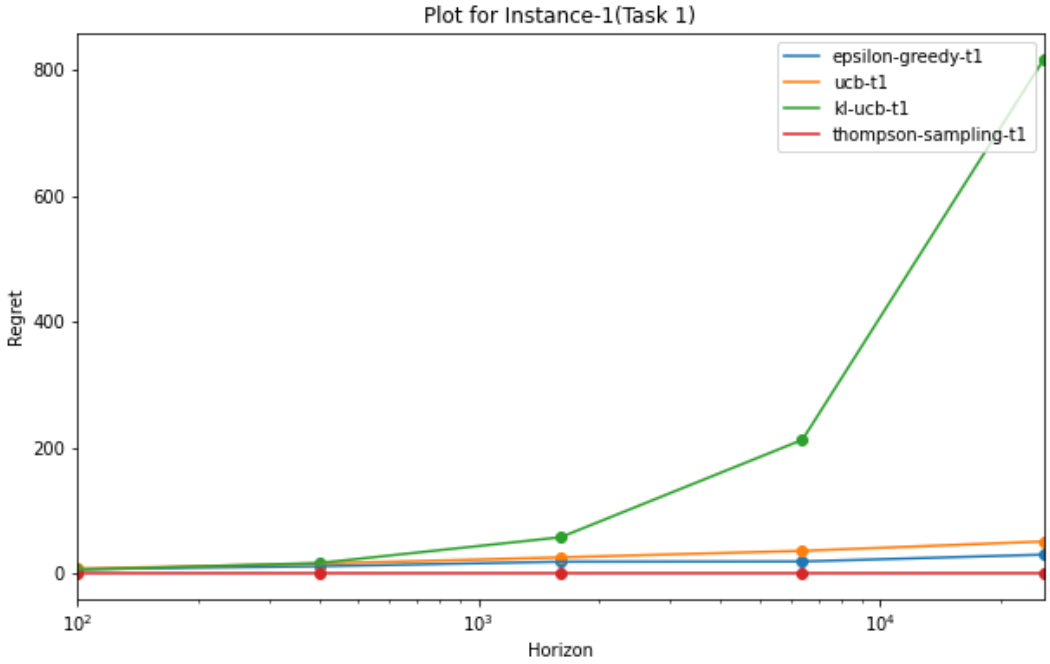
This assignment deals with the implementation of various strategies to minimize regret when pulling a Multi-Armed bandit. Each arm gives a result using a Bernoulli distribution between 0 and 1.

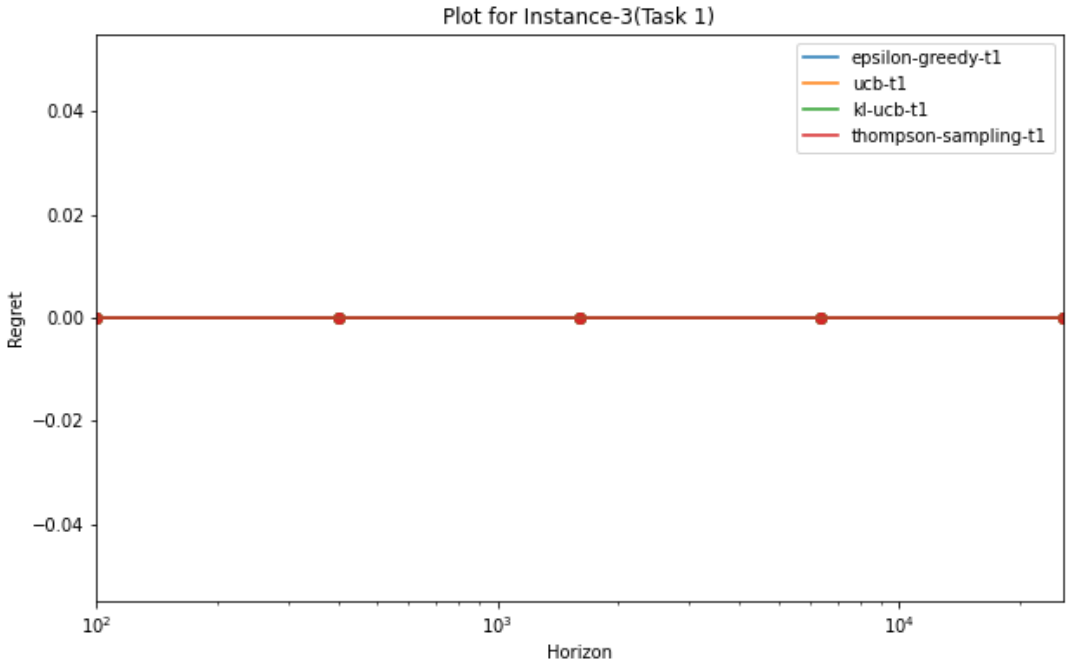
Tasks to be implemented

- **Task 1** (Implementation of basic algorithms like Epsilon-Greedy, UCB, KL-UCB, and Thompson's algorithm.
- **Task 2** (Improving on the implementation of the UCB algorithm by taking a range of c values and finding the optimum of that.
- **Task 3** (Using larger support of rewards instead of a Bernoulli distribution for the above tasks)
- **Task 4** (Maximizing the number of HIGHS for task 3 when the support taken is compared to a threshold value)

Task 1

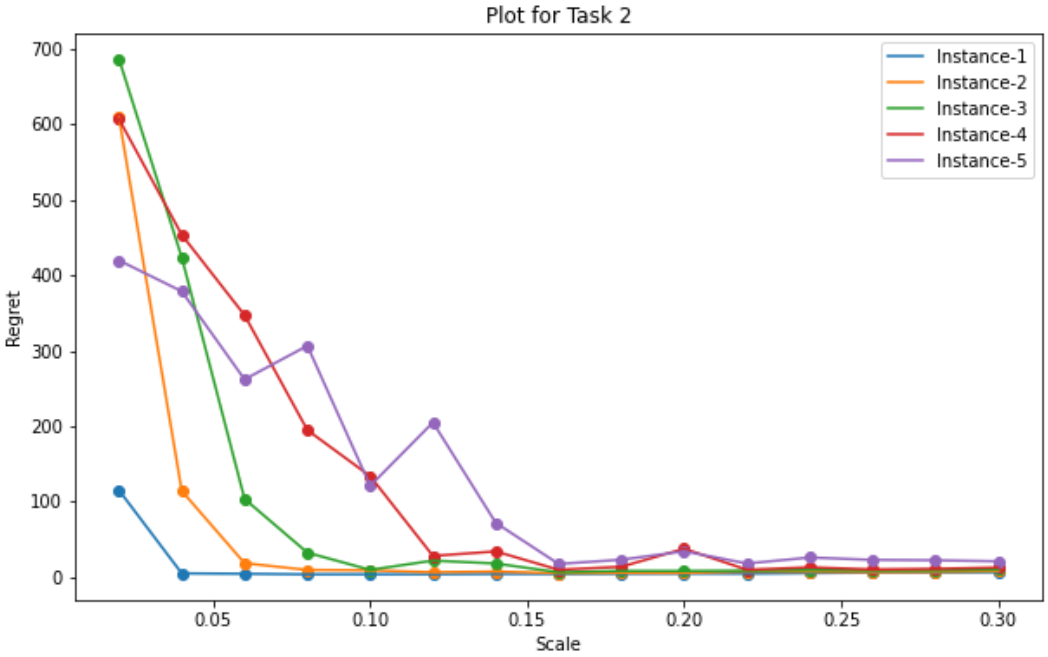






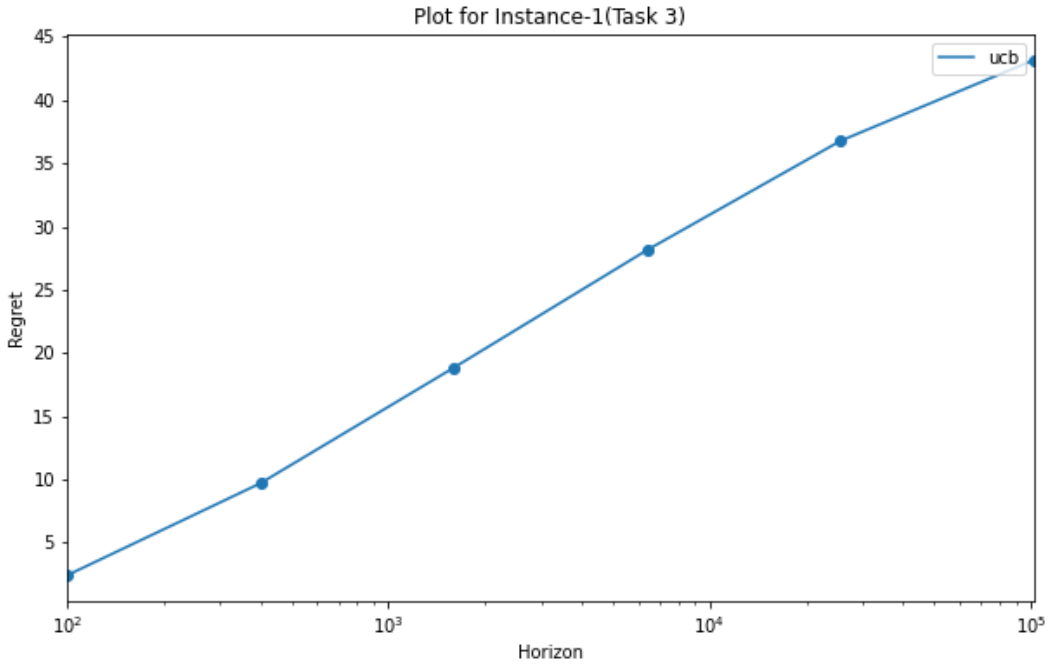
We can infer from the graphs that Thompson's algorithm, the KLUCB algorithm and the UCB algorithm perform better than the Epsilon greedy algorithm (In order). The rate of increase of regret of Thompson's algorithm is much less than the Epsilon-Greedy algorithm.

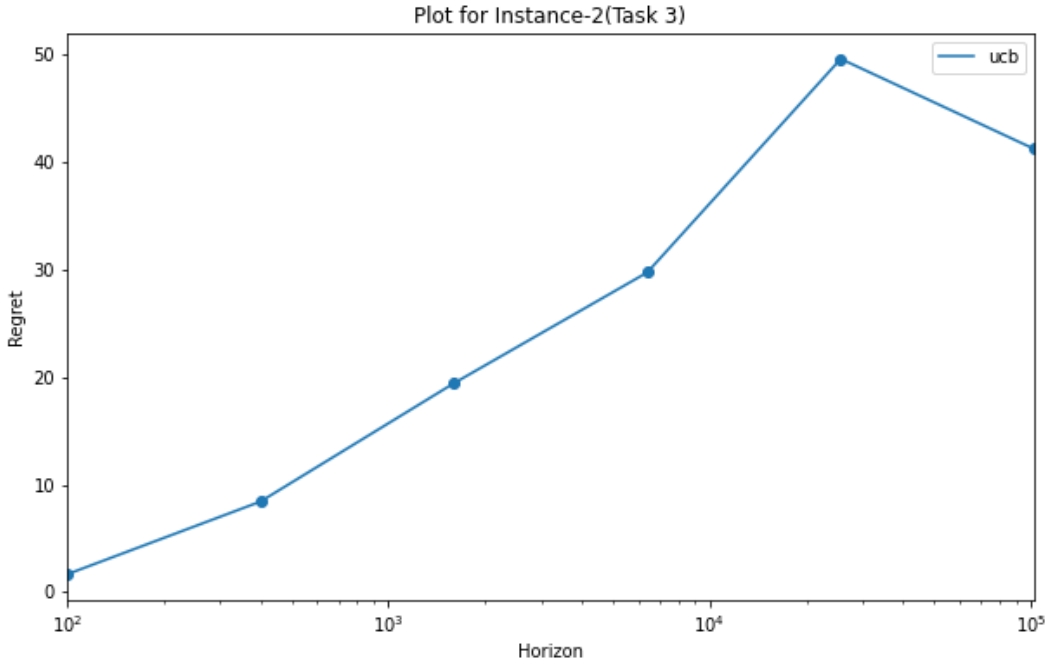
Task 2



From this plot, we can clearly see that changing the value of c greatly affects the performance of the UCB algorithm. The regret drops down most quickly for instance 1 as compared to the other instances.

Task 3





In Task 3, what we have done is, taken Thompson's algorithm, since it has the lowest regret, and tried to optimize that for the given situation. What we have essentially done is, taken the

support values of the bandit's arms and if selected, are directly added to the successes of the arms. This means that the values of successes will not only be composed of 1's but decimals as well since the arm can spew out decimal values as well.

Task 4

For task 4, what we did is, compared the individual output support values of the arms to the threshold, and consequently, output 1 or 0 based on the comparison. The code works perfectly, unfortunately, I was not able to plot a graph as i couldn't generate the data in time and ran into some bugs along the way.