Sheel Shah 19D070052 CS 747 Assignment 1

Task 1 Epsilon Greedy:

1 pull per arm is done initially. The EG-3 algorithm taught in class is then followed.

Task 1 UCB:

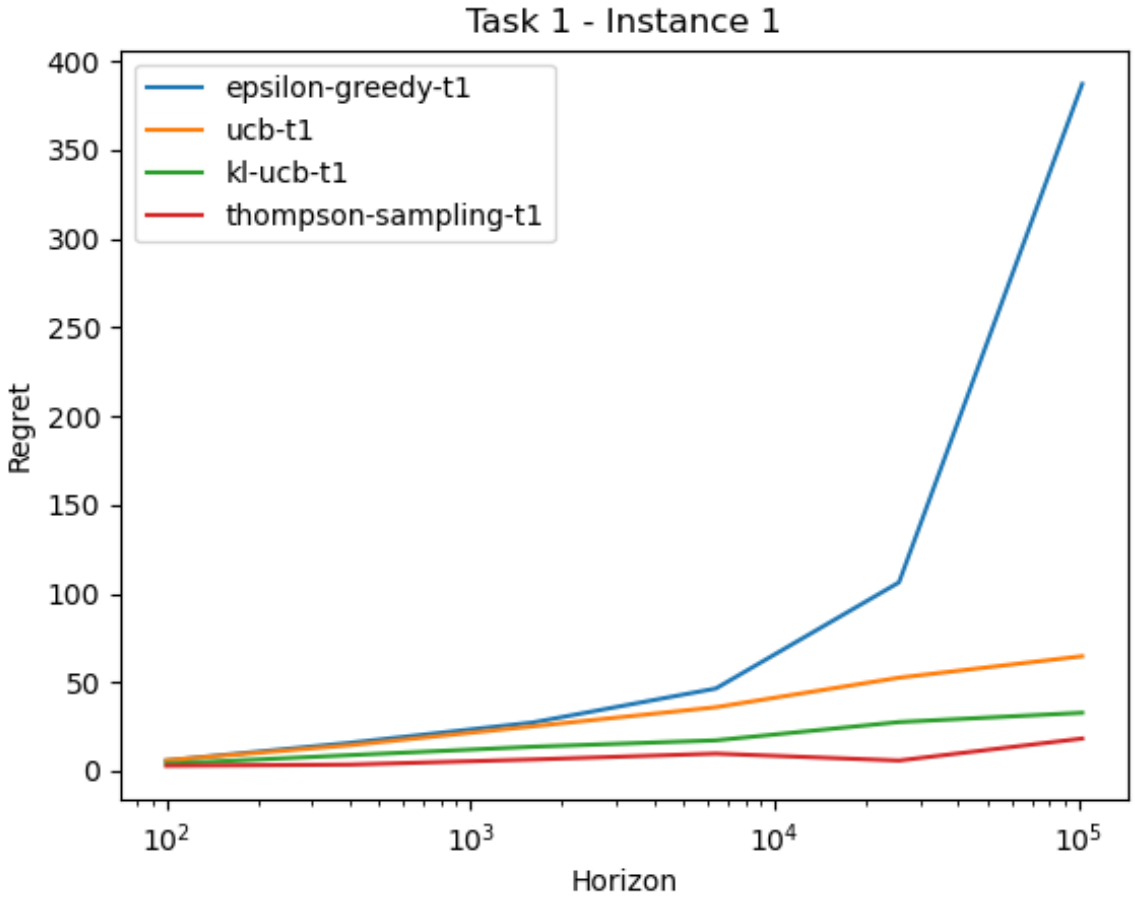
1 pull per arm is done initially. The ucb algorithm is followed with variable scaling factor, which defaults to 2 for task 1.

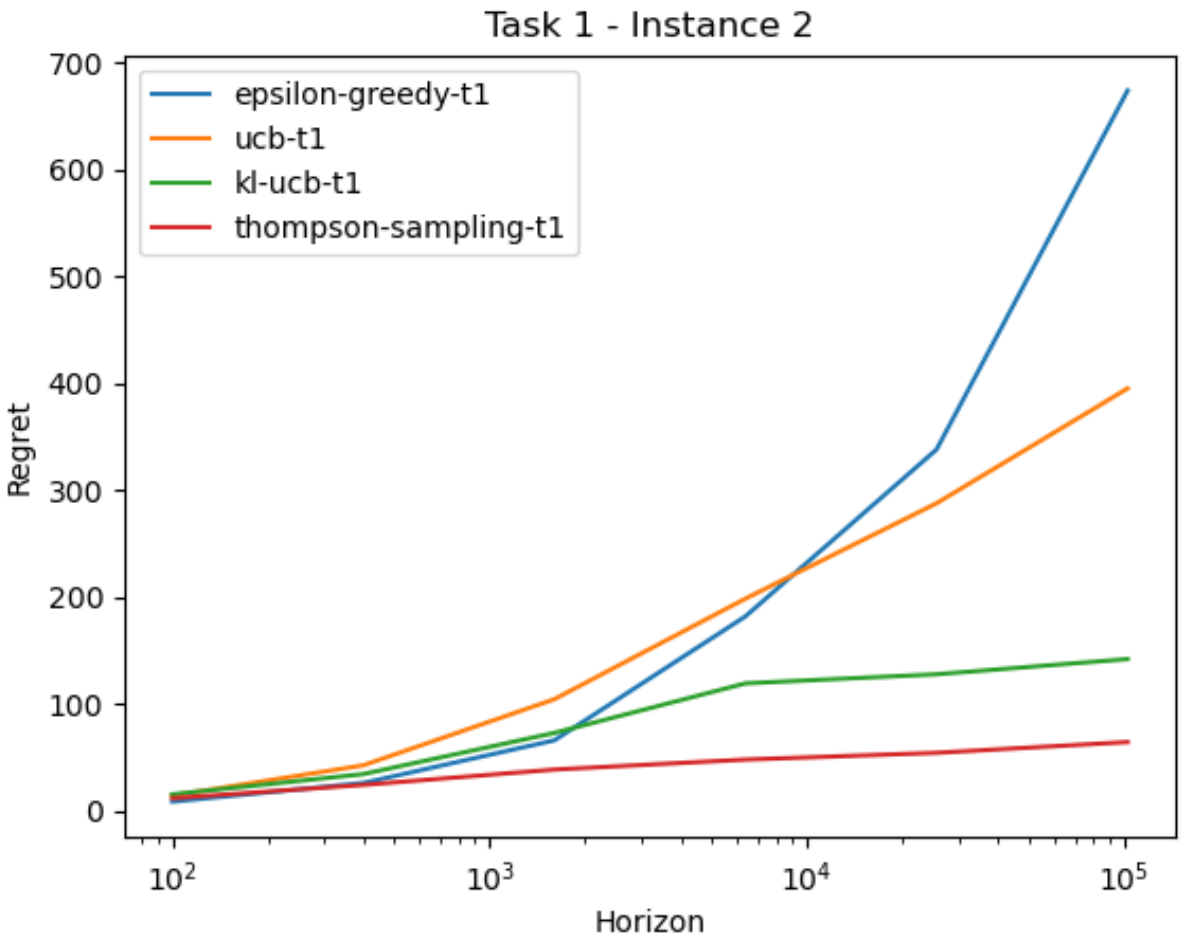
Task 1 KL-UCB:

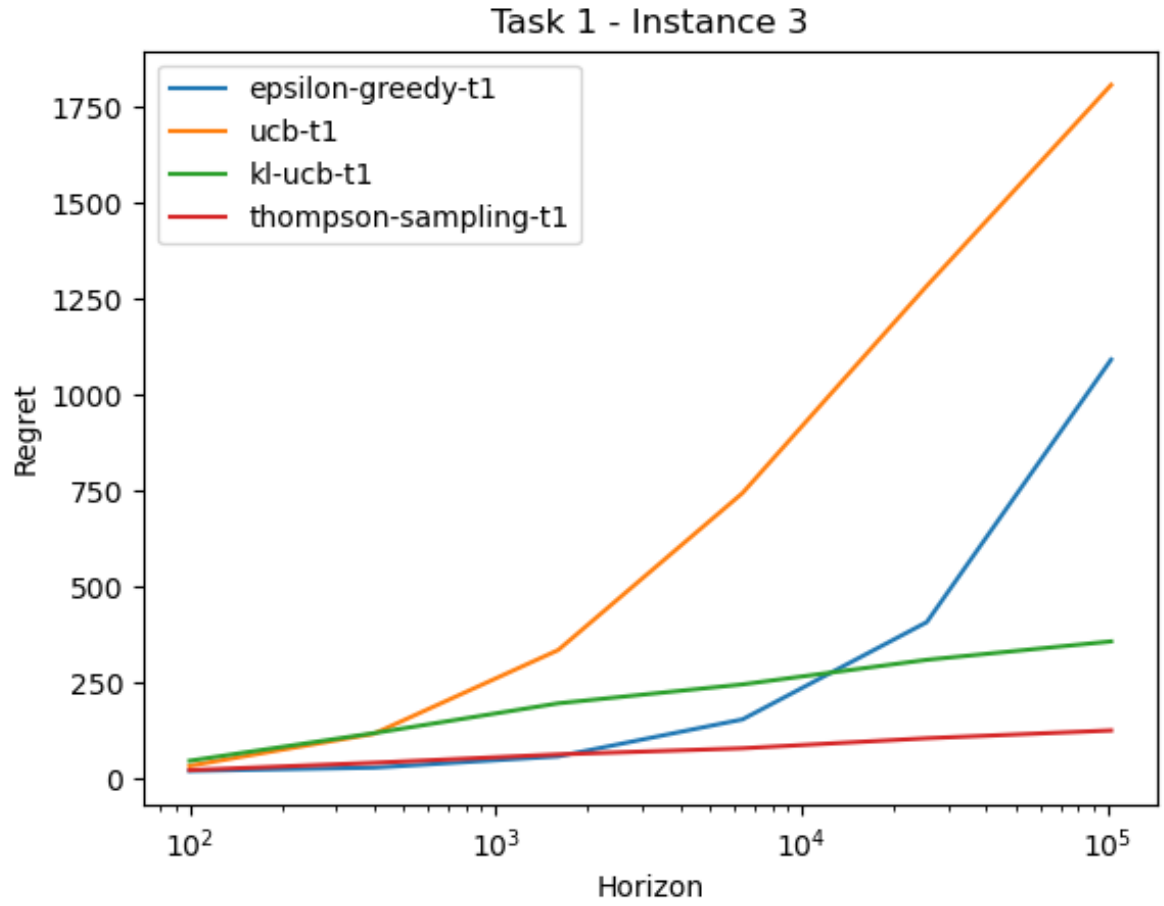
1 pull per arm is done initially. The kl-ucb algorithm is followed with c=3. The q value is calculated programmatically by breaking the necessary interval into steps, and finding the first q for which the kl-ucb bound is dissatisfied.

Task 1 Thompson Sampling:

There are no assumptions. The algorithm is exactly followed as taught in class.




Task 1 - Instance 2
epsilon-greedy-t1
ucb-t1
kl-ucb-t1
thompson-sampling-t1
Regret
Horizon
0
100
200
300
400
500
600
700
10^2
10^3
10^4
10^5

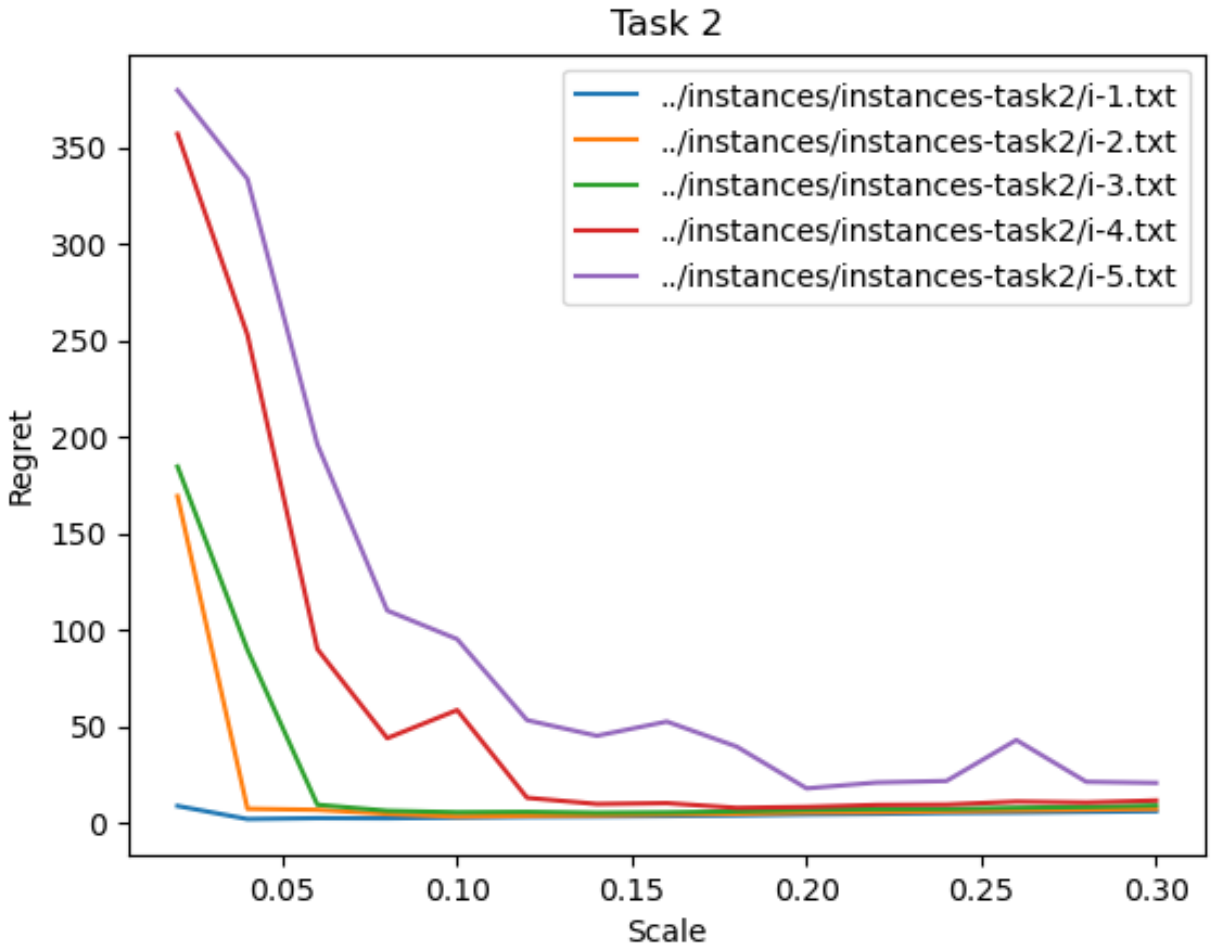

Task 1 - Instance 3
epsilon-greedy-t1
ucb-t1
kl-ucb-t1
thompson-sampling-t1
Regret
Horizon
0
250
500
750
1000
1250
1500
1750
10^2
10^3
10^4
10^5

Task 2 UCB:

The same implementation of task 2 ucb is used, as it was generalized. We see that across the instances, a higher value of c is required for the best case regret. This is because the means of the arms are getting closer, implying we need more exploration in order to be just as sure about the optimal arm.

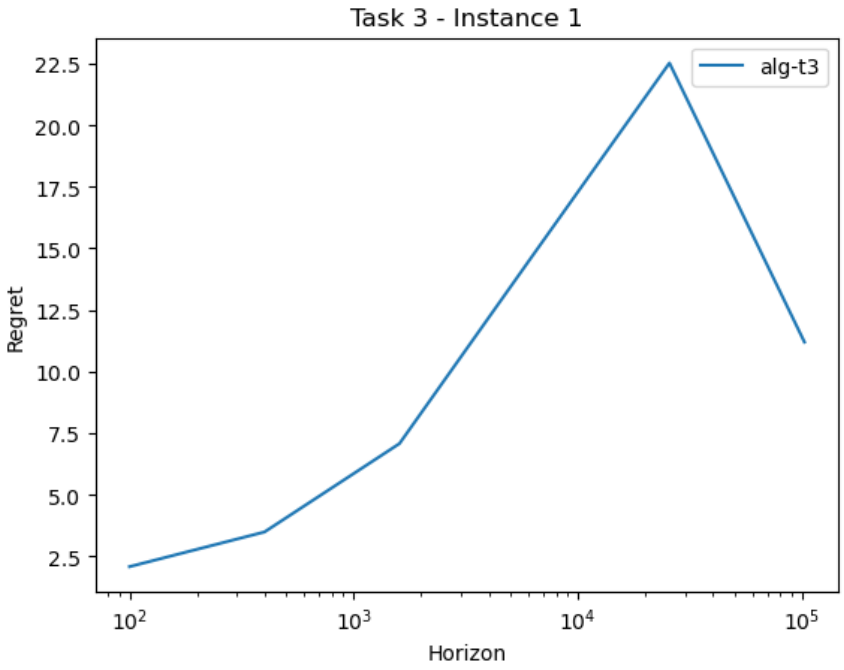
The best scales were found to be:

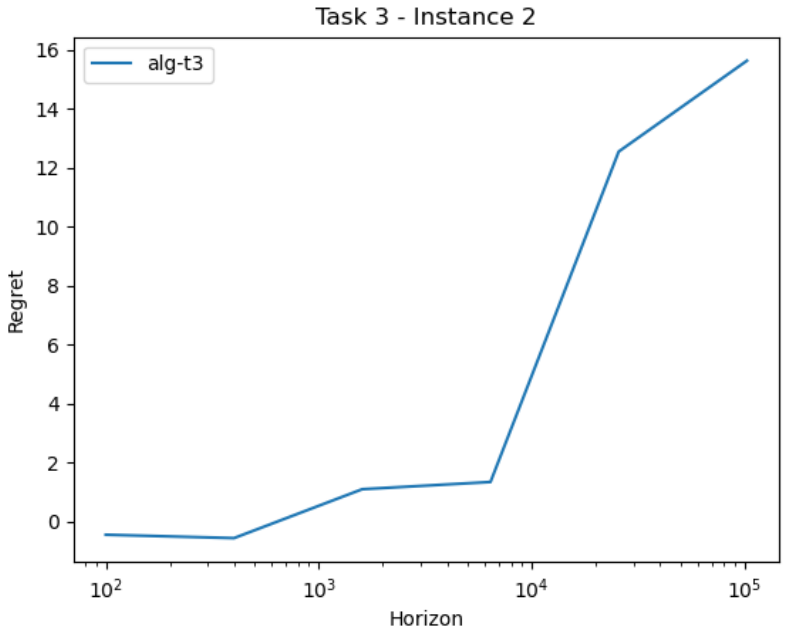
- Instance 1: c = 0.04
- Instance 2: c = 0.1
- Instance 3: c = 0.14
- Instance 4: c = 0.18
- Instance 5: c = 0.2



Task 3 Algorithm:

I looked up the Dirichlet distribution which extended the beta distribution to multinomial variables. A sample from the Dirichlet distribution is drawn, where the alphas are the number of times each reward in the support is seen + 1. (similar to Thompson sampling). This sampled vector is dotted with the support of this arm, to create a sample_value. This approach should work since the expectation of the sample_value is the same as the current expected reward of this arm, and the variance dies down with more and more pulls (just like Thompson sampling). It should be noted that the support had repetitions which were eliminated by only considering unique values of the support.





Task 4 Algorithm:

This task can be mapped to the Bernoulli case. A reward is 1 if it is greater than the threshold and 0 otherwise. Hence the support can be split at the threshold, and the probabilities cumulated to map the multinomial arm distribution to Bernoulli. Now, Thompson sampling can be directly followed where a success is when reward > threshold.