CS747 Assignment 1
Manoj Bhadu(170010036)

T1:

Assumptions:

1. **For Epsilon-Greedy:**
    a. First number of arms time steps exploration for breaking tie
    b. Epsilon value = 0.02
2. **UCB:**
    a. First number of arms time steps pull every arm so that zero division does not occur in second term
    b. Then the number of arms time steps exploration for breaking ties.
3. **KL-UCB:**
    a. First number of arms time steps pull every arm so that zero division does not occur in second term
    b. Used binary search for finding maximum q and used precision of 1e-6
    c. Used C=0

4. **Thompson sampling:**
    a. No exploration, start thompson sampling direct because there is no problem of zero division
    b. Used python random.betavariate module for sampling

T2:

Approach:

1. I tried to match the max of empirical means to the max of true means of arms. So at every timestep I checked if my max of empirical mean of arms is greater than the second max of true mean or not. If yes sample the arm with maximum empirical mean and if not sample according to beta distribution
2. As we reach close to the optimal arm the algorithm will sample that arm only assuming we are going in the right direction. For a longer horizon and more number of arms the max of empirical mean converges to the mean of the optimal arm.
3. At first time step sampling is done according to the beta distribution.

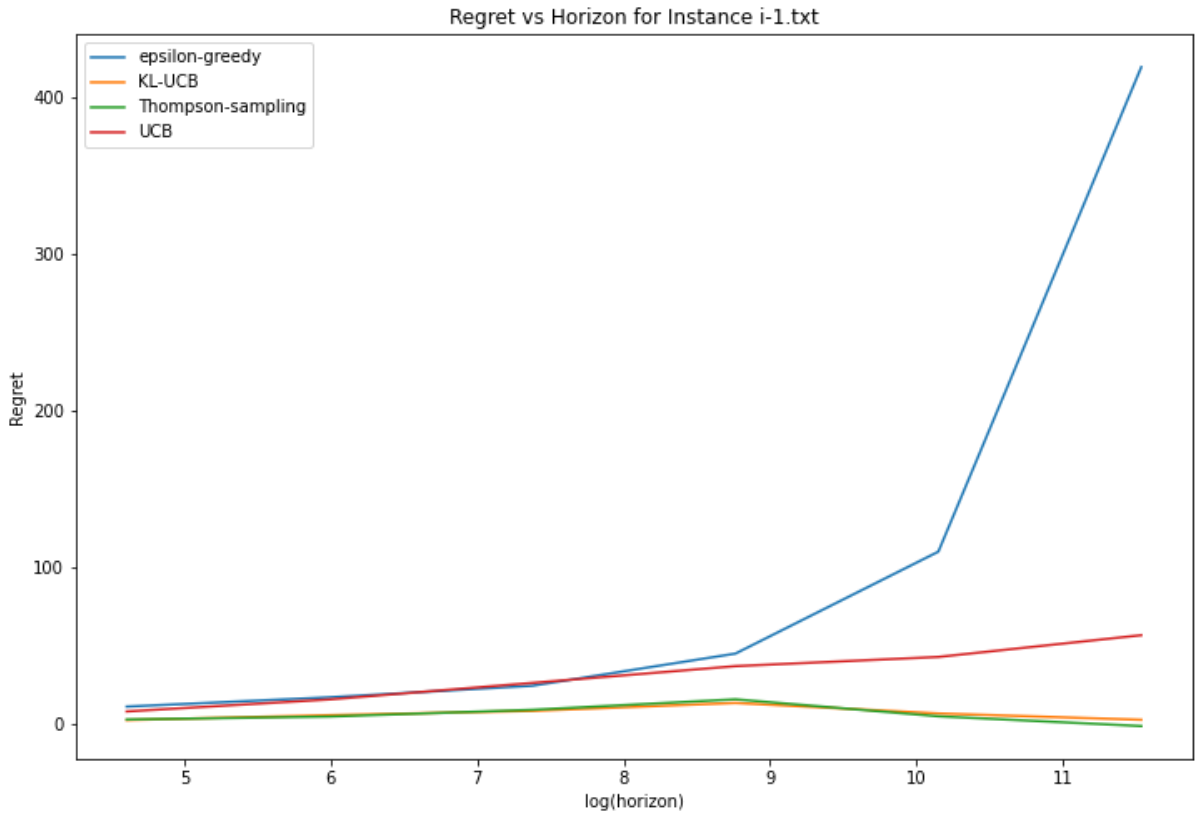
T3:

1. For instance i-1 value of $\epsilon_1 = 0.0001, \epsilon_2 = 0.100001, \epsilon_3 = 0.200001$
2. For instance i-2 value of $\epsilon_1 = 0.001, \epsilon_2 = 0.100001, \epsilon_3 = 0.200001$
3. For instance i-3 value of $\epsilon_1 = 0.000001, \epsilon_2 = 0.100001, \epsilon_3 = 0.200001$

T4:

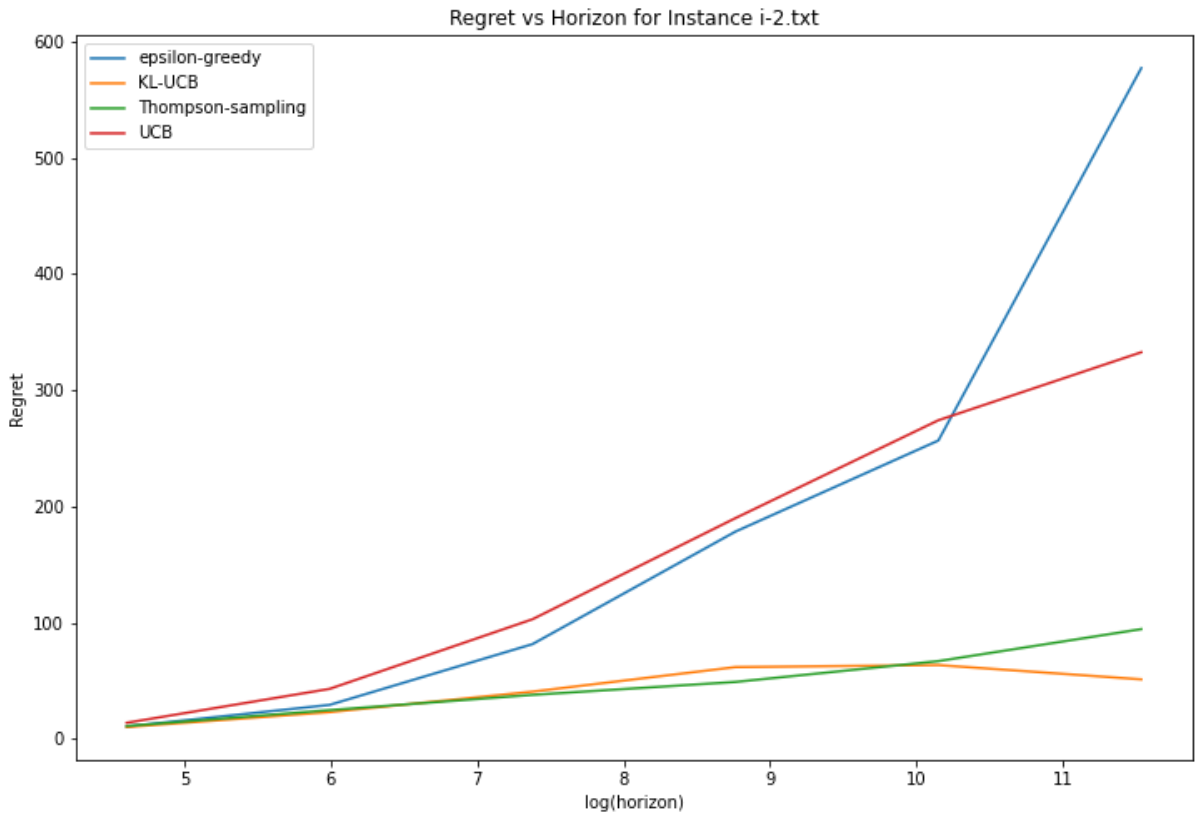
Plots and Results:

1. **Instance i-1**
    a. **Plot**



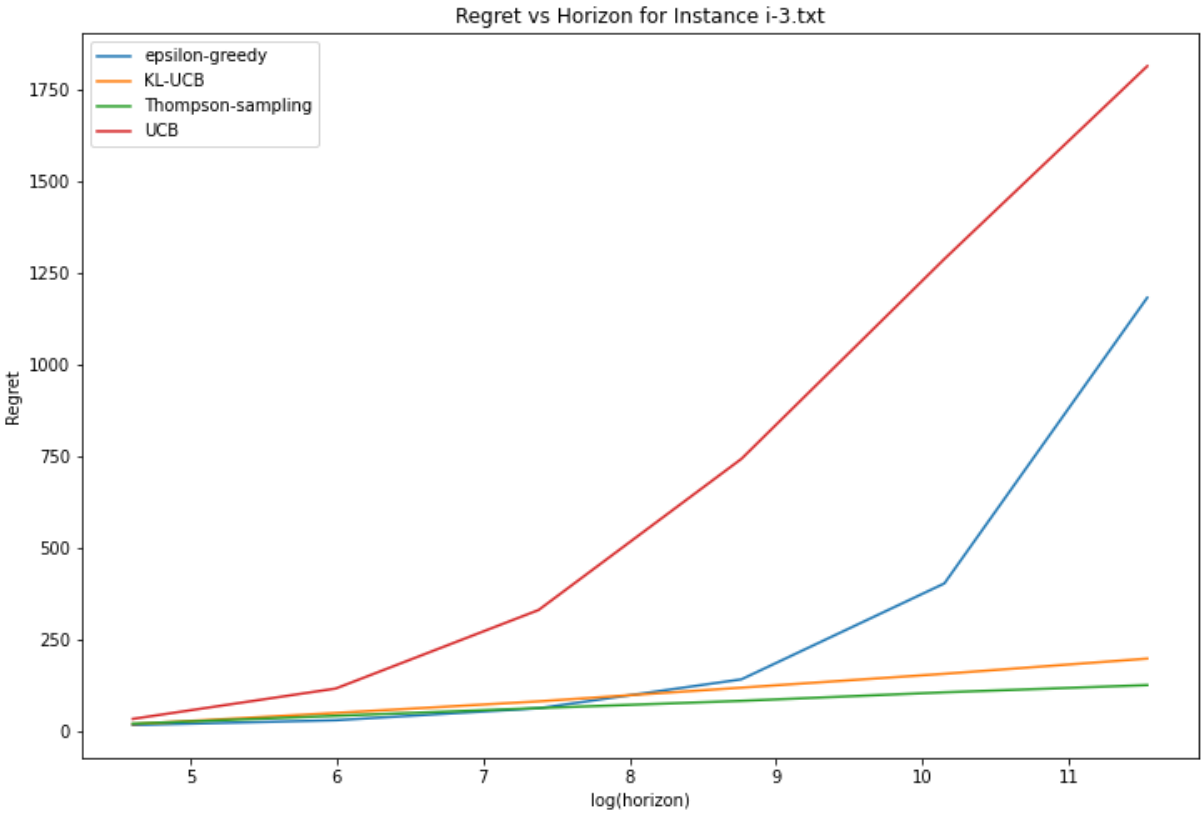
    b. **Comments**: As we can thompson-sampling and KL-UCB gives the minimum regret which is expected and regret is also approx some constant multiple of log of horizon for both. UCB is performing better than epsilon greedy here.

2. **Instance i-2**
    a. **Plot**



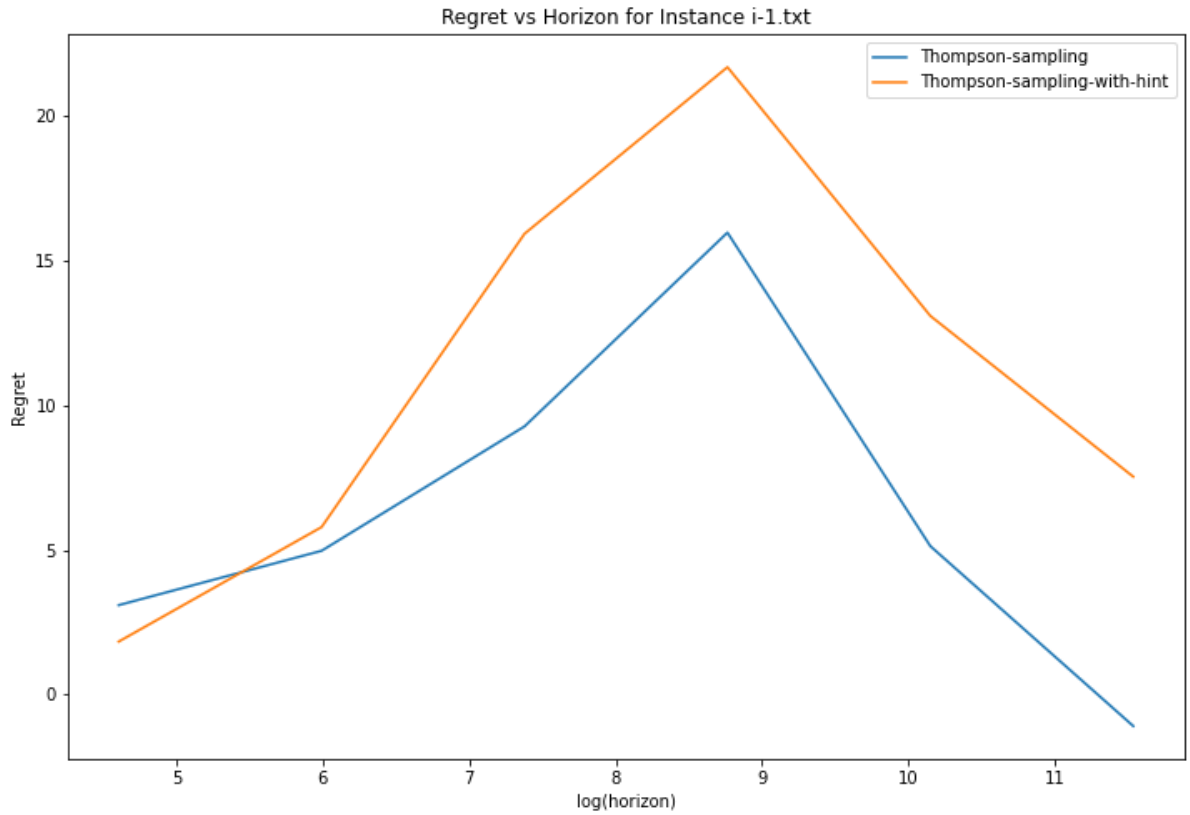
b. **Comments:** Here also we can see the thompson-sampling and KL-UCB gives the minimum regret and KL-UCB performing better at larger horizons and Regret is some constant time multiple of log of horizons. Epsilon greedy and UCB giving similar results initially but for larger horizons UCB start giving better performance

3. **Instance i-3**
    a. **Plot**



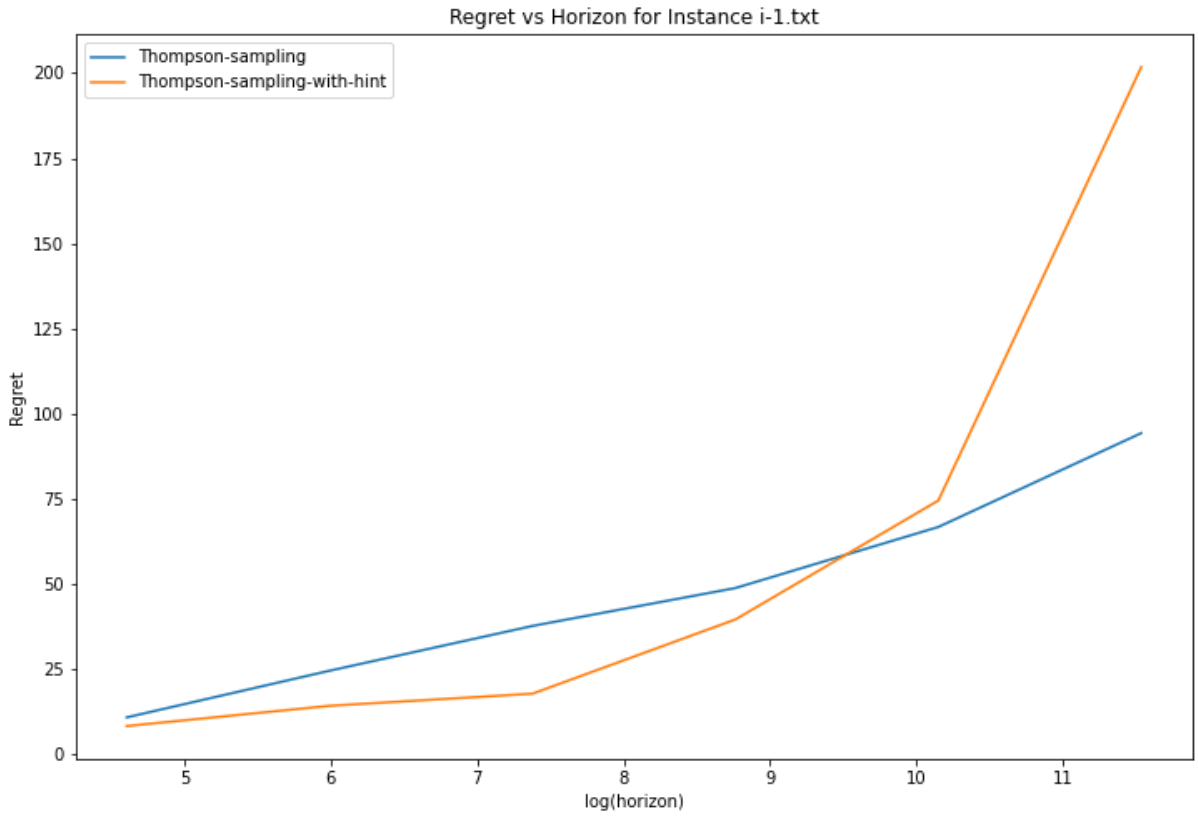
b. **Comments:** Here also we can see the thompson-sampling and KL-UCB gives the minimum regret and KL-UCB performing better at larger horizons and Regret is some constant time multiple of log of horizon. Here epsilon-greedy performs better than UCB which is unexpected. Maybe for some seeds UCB gives larger regret which increases the regret. But here from data we can see UCB gives the larger regret from 102400 horizon from epsilon greedy most of time.

4. **Instance i-1 for thompson sampling with hint:**
    a. **Plot:**



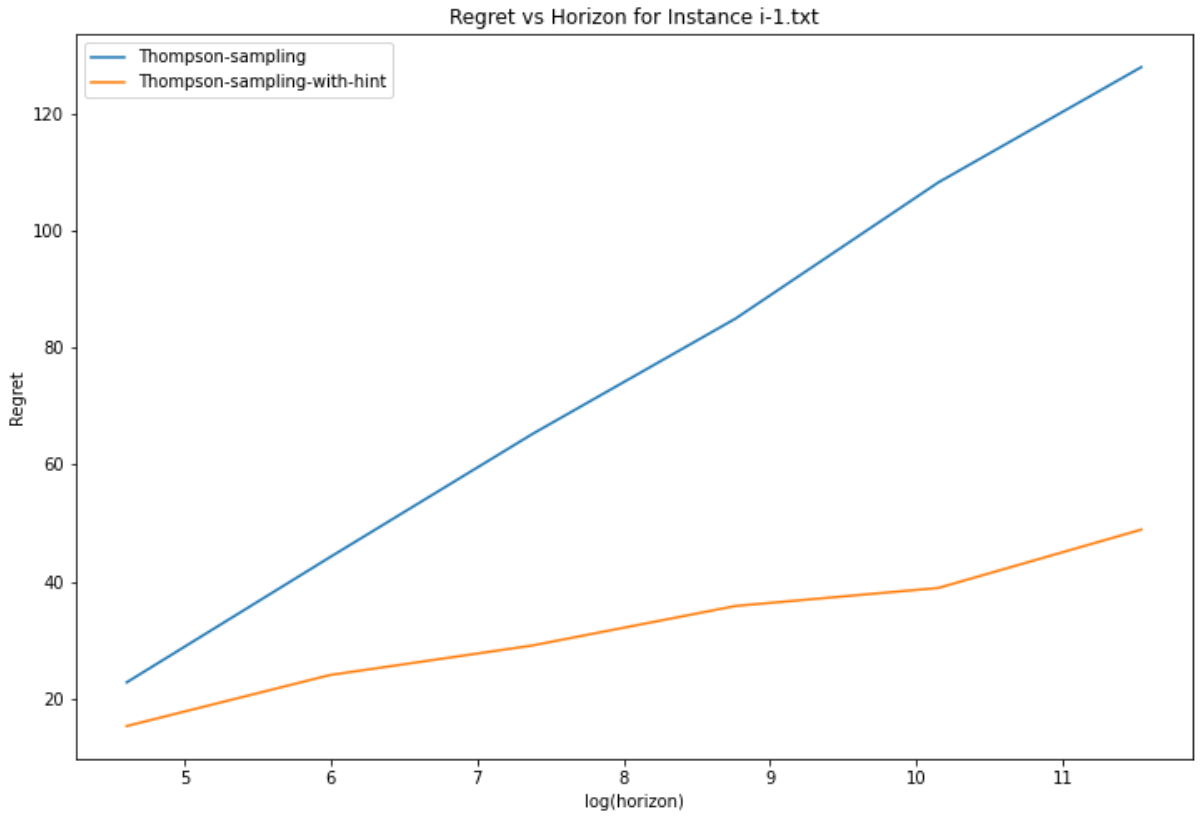
    b. **Comments:** As we can see thompson sampling with hints not performing well than normal thompson sampling because the number of arms are less and probably we may have assumed the second arm is optimal and pulling that arm.

5. **Instance i-2 for thompson sampling with hint**
    a. **Plot:**



    b. **Comments:** We can see that initially thompson sampling with hint performing better than thompson sampling but for larger horizon not performing well because we are constantly exploring the same arm whose empirical mean converged to max of true mean of arms and that arm also produce 0 reward with some probability. This is kind of a subjective discussion but that's my interpretation.

6. **Instance i-3 for thompson sampling with hint**
    a. **Plot**



    b. **Comments:** As expected for more numbers of arms our Thompson sampling with hint performs better and the difference is increasing for larger horizons. These results are expected and the optimal arm is what we are sampling after the max of the empirical mean converges to the optimal arm's true mean.