Technical University of Crete

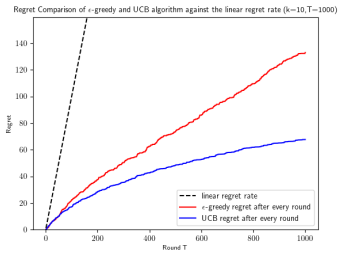
School of Electrical and Computer Engineering

Course: **Reinforcement Learning and Dynamic Optimization**

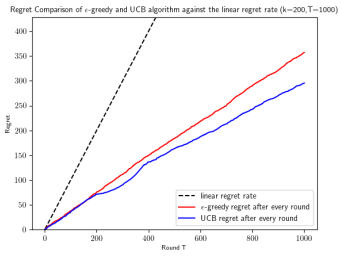
Assignment 1

Report Delivery Date: Sunday, March 26, 2023

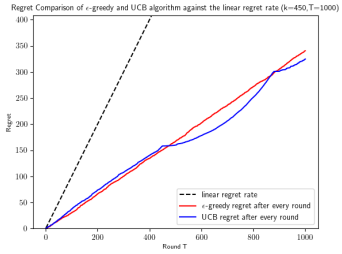
Student: Alevrakis Dimitrios 2017030001



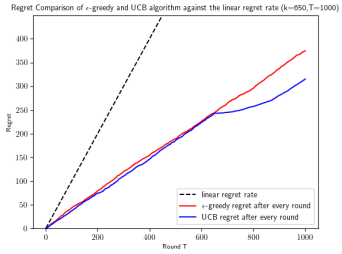
(a) $(k = 10, T = 1000)$



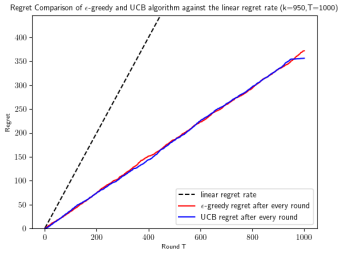
(b) $(k = 200, T = 1000)$



(c) $(k = 450, T = 1000)$



(d) $(k = 650, T = 1000)$



(e) $(k = 950, T = 1000)$

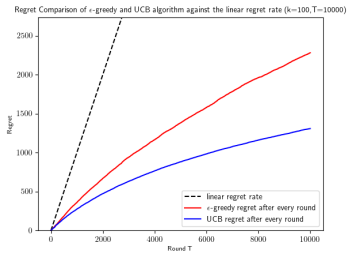
Figure 1: $(k \in [10, 950], T = 1000)$

1. In the experiments of figure 1, we are testing the effect that the ratio $\frac{k}{T}$ has on the performance of both algorithms.
   We can observe that, when $k \ll T$ the UCB algorithms achieves a distinguishably lesser regret rate and thus lower regret at the horizon.
   Though, when $k < \frac{1}{2}T$ the UCB algorithm has slightly lower regret rate and achieves

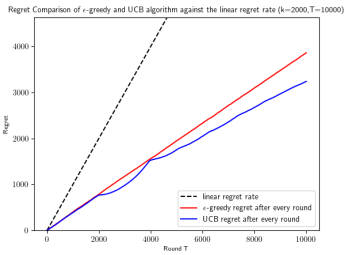
a lower regret at the horizon but by a significantly lesser margin.

When k can be considered $k \approxeq \frac{1}{2}T$ the algorithms have similar regret rates. Worth noting that UCB presents "spikes" where the regret becomes lower than t ϵ-greedy but after again equalizing as shown in figure 1 (c) and (d). Those spikes are probably due to the algorithm choosing the best hand but afterwards switching.

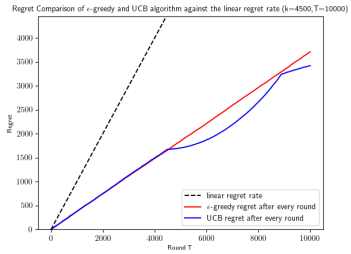
Lastly when $k \approx T$ both algorithms present the same regret rate.



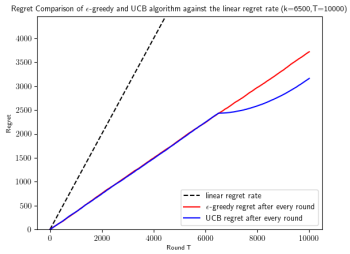
(a) $(k = 100, T = 10000)$



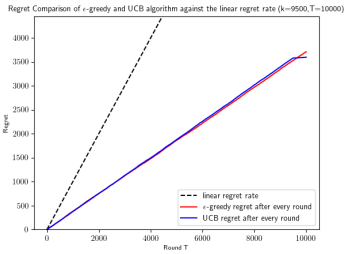
(b) $(k = 2000, T = 10000)$



(c) $(k = 4500, T = 10000)$



(d) $(k = 6500, T = 10000)$



(e) $(k = 9500, T = 10000)$

Figure 2: $(k \in [100, 9500], T = 10000)$

The experiments of figure 1 where tested again for different values of k, T but with the same $\frac{k}{T}$ ratios. As shown in figure 2 the algorithms present the same behaviors observed in figure 1.