IEMS-469 Dynamic Programming HW4

Weijia Zhao [1]
Kellogg Finance

This version: December 7, 2021

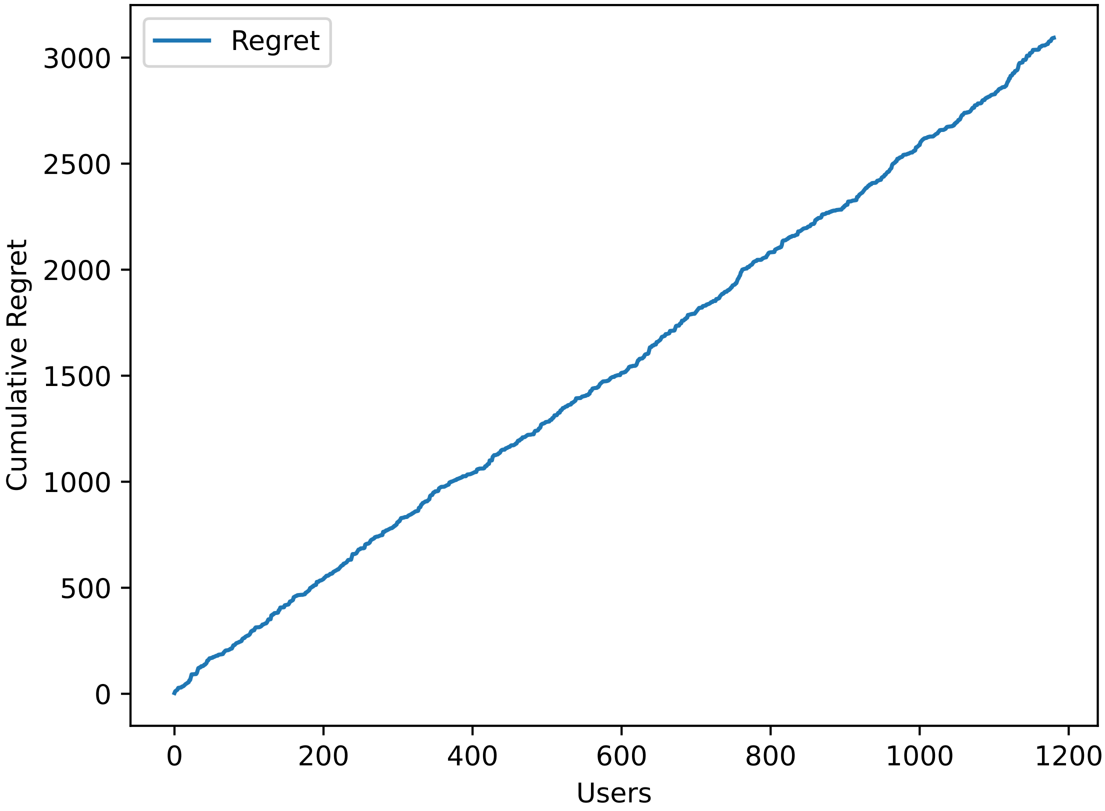
[1] Weijia.Zhao@kellogg.northwestern.edu

1 Jester/UCB

I set the weight of bonus to be $\alpha = 1$. I disable the update of parameters when use the model in ”testing” mode (i.e. this is different from online learning, when we are given new samples one by one and update the parameters in the meantime. In my analysis, the actual reward can only be observed altogether in the end when doing evaluation).

Seems to me that we can choose to go over the training sample multiple times but the improvement of performance is not very significant. In general, the cumulative reward is very close to a linear function of the number of observations in testing set. The total regret for 1200 observations is about 3000 (I did several experiments and the highest I got is 3091 and the lowest I got is 3007)

Figure 1: Jester


Regret
Cumulative Regret
Users