FILA Assignment 1

Section 1) Changes due to new form of continuous reward function

Main change was made in Thompson sampling:
The success and failure calculation was done by adding the reward for success and adding (1-reward) for failure

Section 2) Implementation of Algorithm

For thompson sampling I generated the beta distribution number by using gamma distribution. It is because of the following relation

Let $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Gamma}(\beta, 1)$ where the paramaterization is such that α is the shape parameter. Then

$$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta).$$

For KL_UCB :

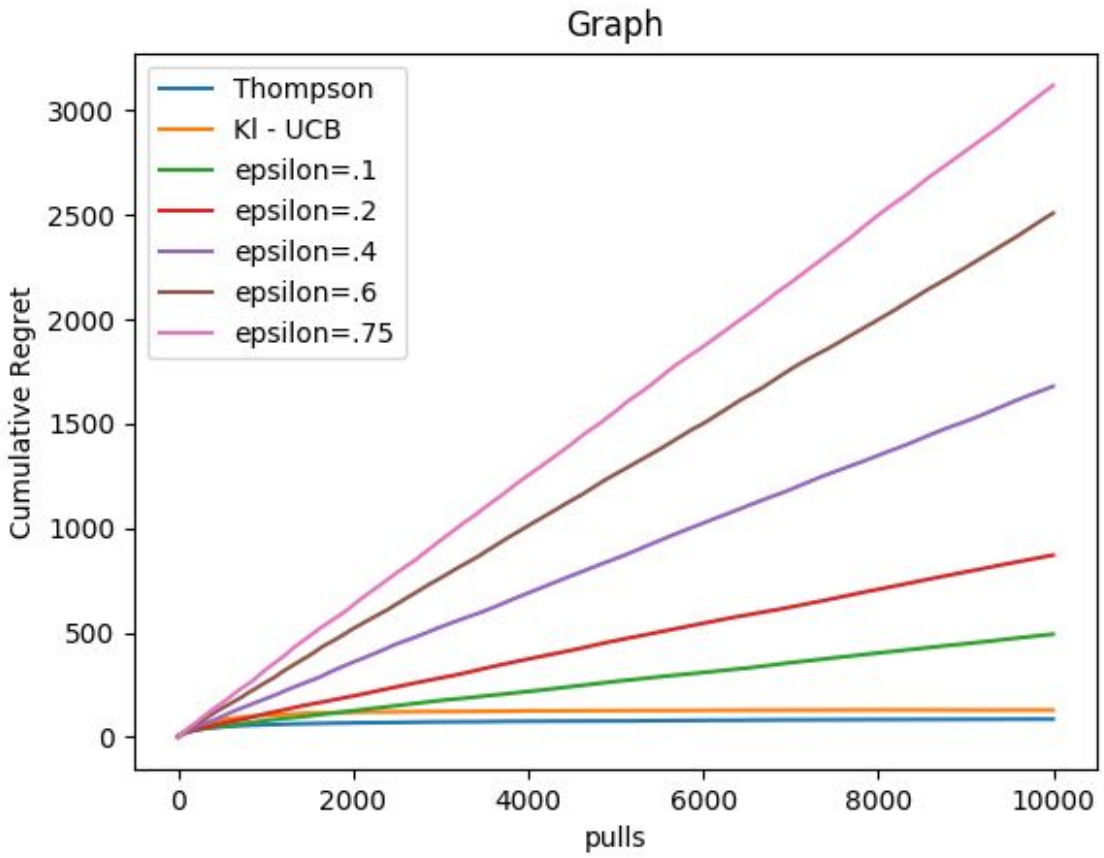
I am regressing towards the “q” value maximum by making steps of size .02 towards 1 and checking if the value is still less then the given ones.

For epsilon-greedy:
I used gsl to generate a random number.

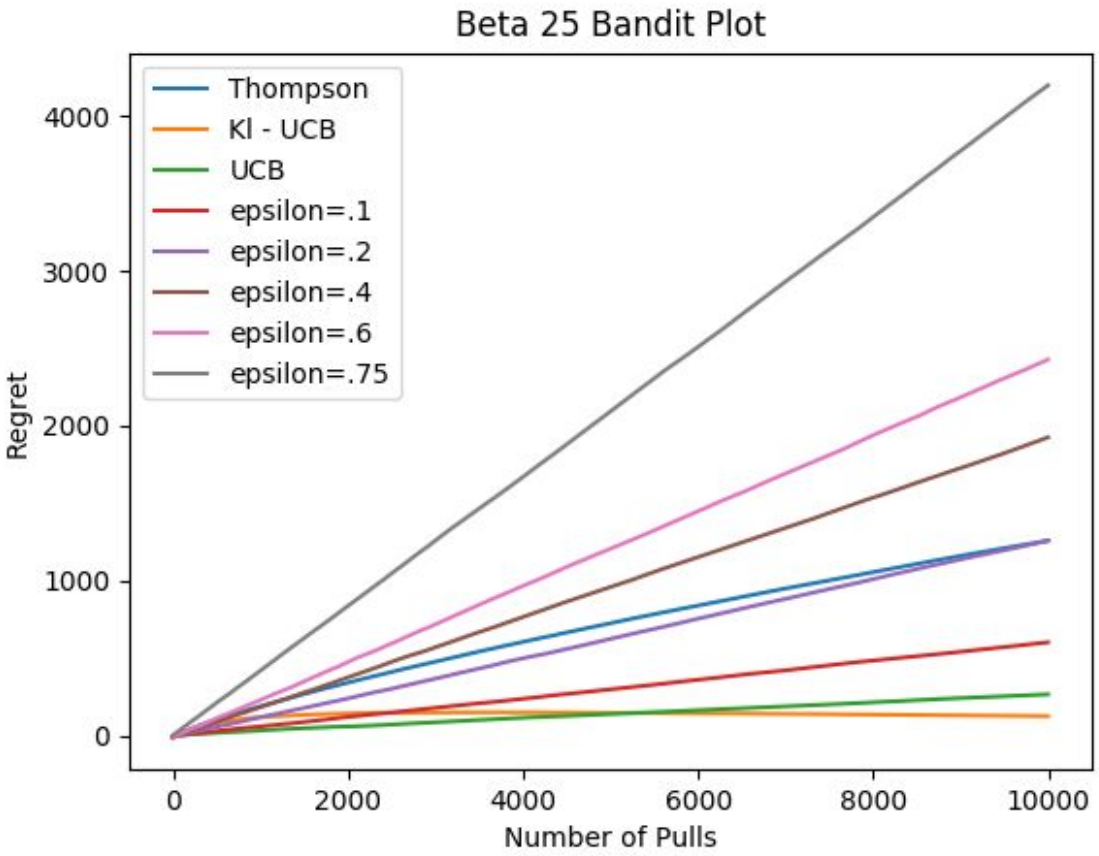
Section 3) Observations

The graphs obtained were as expected for most of the outputs.

a) instance-bernoulli-25



b) betaDist_25

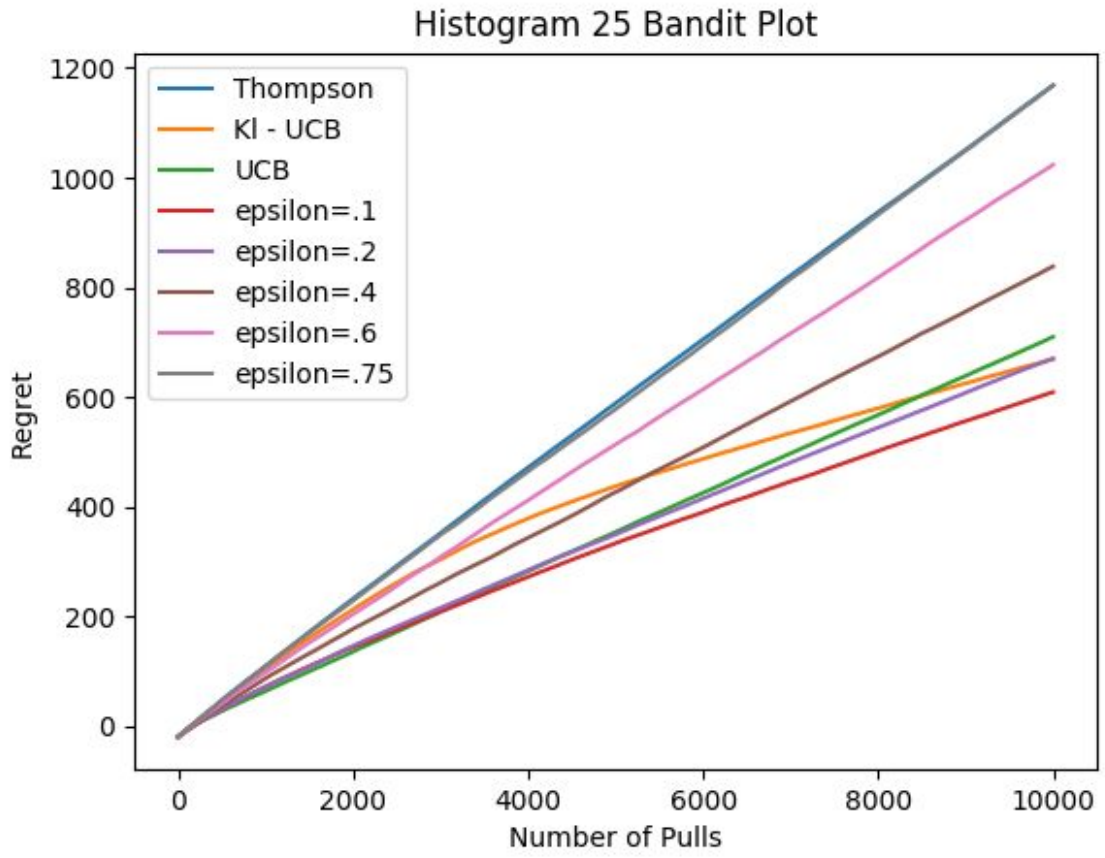


I see some abnormal results in this
The thompson sampling is not performing that well. But it tends towards saturation as the number of pulls increase as we can see in the graph. In fact if we further increase the horizon, thompson sampling will perform better than others

Average Regret values

Thompson [[1257.39822838]
KL UCB [124.42121733]
UCB [265.10256949]
Epsilon .1 [601.26519125]
Epsilon .2 [1257.27232154]
Epsilon .4 [1924.53628453]
Epsilon .6 [2426.94555448]
Epsilon .7 [4197.59]]

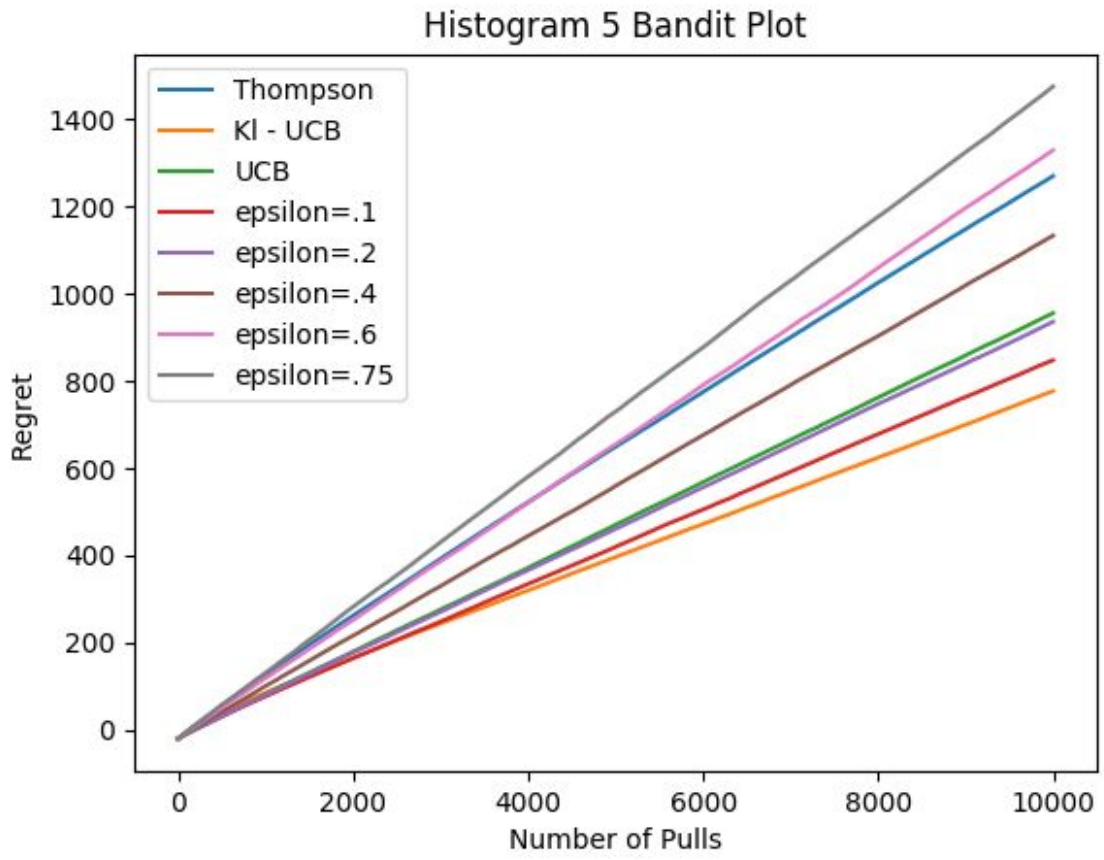
c) instance-histogram-25



Again in this we see that the KL UCB and Thompson are initially performing bad, infact lower than epsilon .1, but both are tending to saturated regrets, and hence if we increse the pulls the results will turn to be as expected.

Average Regret values
Thompson [[2667.97899435]
KL UCB [2168.9573535]
UCB [2209.4906103]
Epsilon .1 [2108.65497967]
Epsilon .2 [2169.78896686]
Epsilon .4 [2337.94391091]
Epsilon .6 [2523.3511309]
Epsilon .7 [2668.27422968]]

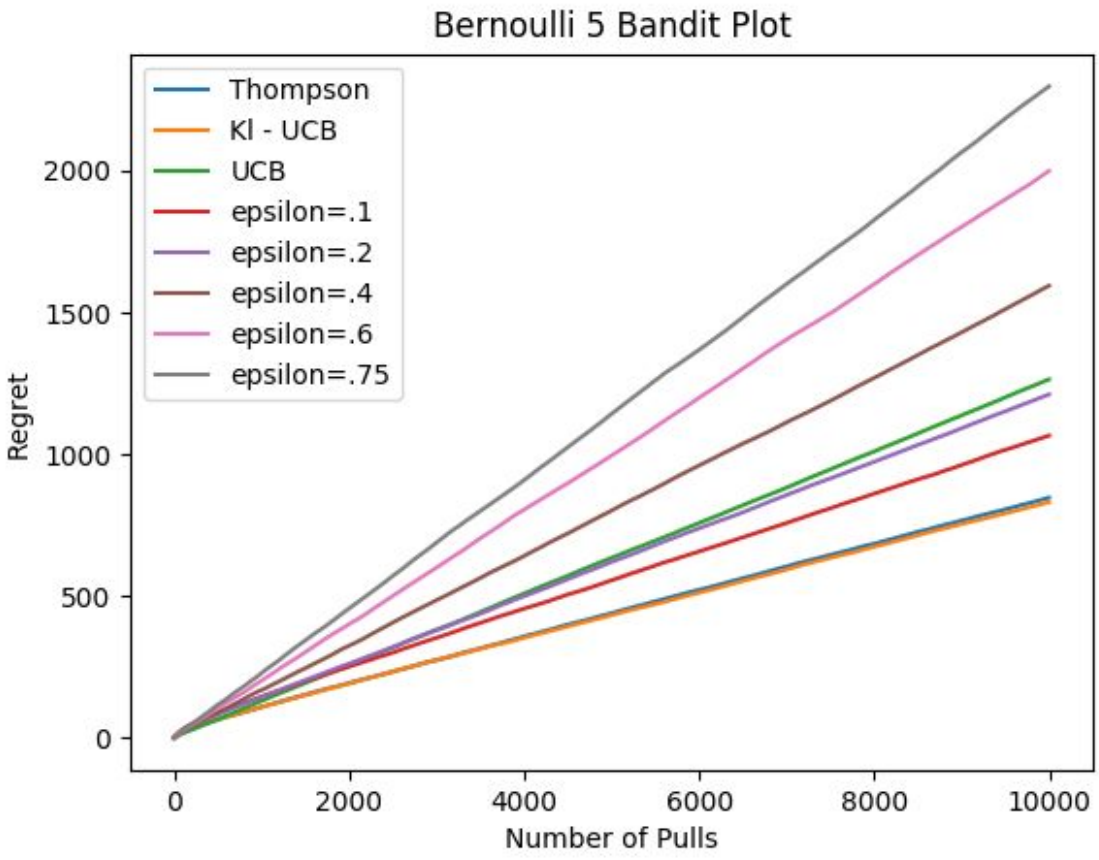
d) instance-histogram-5



Kl UCB performs well

Thompson [[1269.89295431]
KL UCB [776.86482428]
UCB [955.66877085]
Epsilon .1 [847.73803954]
Epsilon .2 [935.87970554]
Epsilon .4 [1133.23631197]
Epsilon .6 [1329.04176818]
Epsilon .7 [1475.12440316]]

e) instance-bernoulli-5



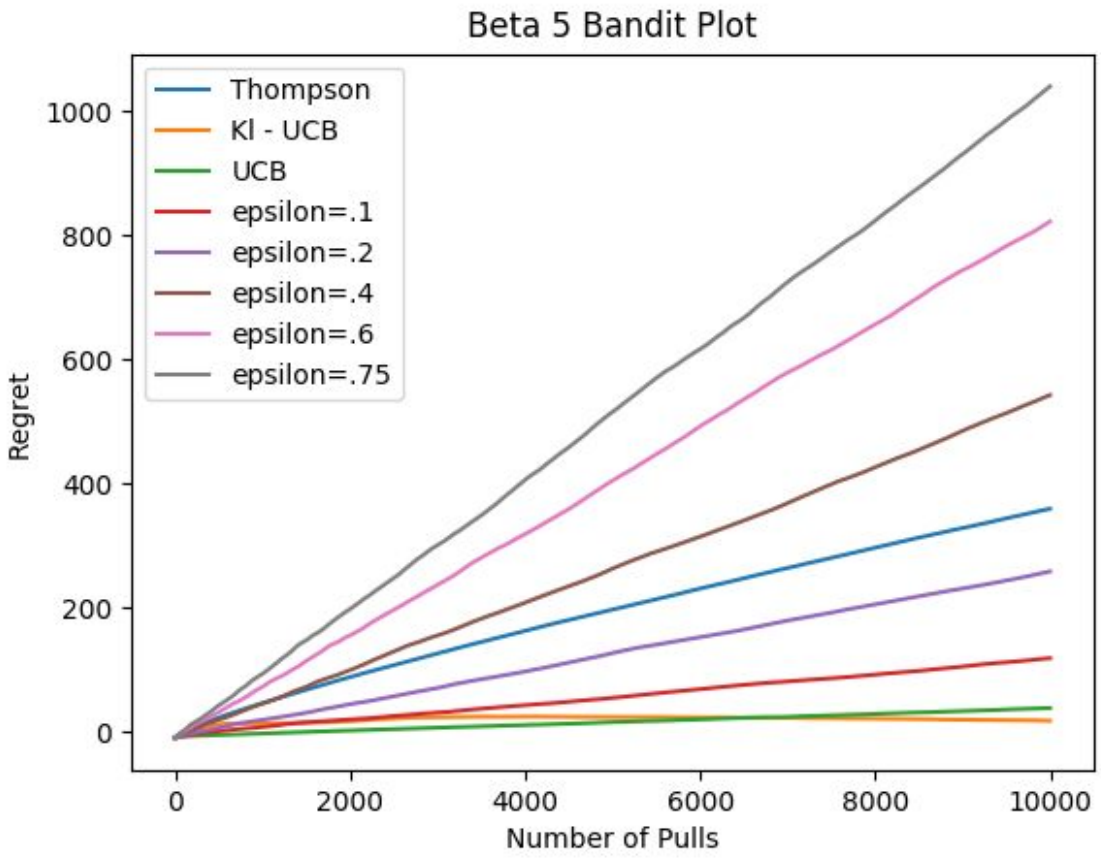
All Algorithms perform as expected, UCB deviates though

Average regrets

Thompson [[847.17]

KL UCB [831.36]
UCB [1264.38]
Epsilon .1 [1066.13]
Epsilon .2 [1211.36]
Epsilon .4 [1594.72]
Epsilon .6 [1999.36]
Epsilon .7 [2297.78]]

F) betaDist_5



Average Values

[[359.29453565]
[17.36463117]
[37.82578365]
[118.20874881]
[258.17483354]
[542.63845871]
[822.61802772]
[1040.60870284]]