CSCI 5525: Homework #5

Noah Hendrickson

Problem #1a:

CSCI 5525: HW5

Noah Hendrickson

Problem #1a:

We consider the class of threshold functions

$H = \{1[x > a] \mid a \in R\}$

where x is a feature and $1[x>a] = 1$ if $x > a$, 0 else.

To find the VCD, we consider a graphic proof. This problem can be represented as a 1D number line going left → right negative → positive.

First we consider the case with one sample: x_1

Case x_1 label $= 0$: $-$ —— x_1 —— a —— $+$

Case x_1 label $= 1$: $-$ —— a —— x_1 —— $+$

In both cases, it is trivial to see that no matter where x_1 is placed, a classifies it as 0 if placed to the right and 1 if placed to the left, thus $VCD(H) \geq 1$.

Second, we consider samples x_1 and x_2.

Case x_1 label $= 0$, x_2 label $= 1$

$-$ —— x_2 —— x_1 —— $+$?

We can see from the left graphic that, when x_1 and x_2 with the provided labels are placed in this configuration, there is no a that perfectly classifies both. If a were placed to the left of x_2, x_1 would be misclassified. Placing a to right of x_1 would misclassify x_2. Inbetween would misclassify both. Thus, $VCD(H) < 2$.

Thus, $VCD(H) = 1$

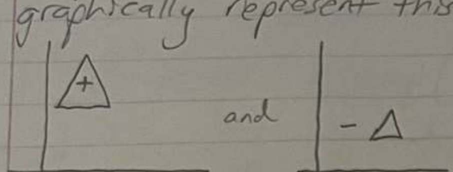
Q.E.D.

Assumption: Both answers to a and b assume 2 data points in the same location in space will have the same label.

Problem #1b:

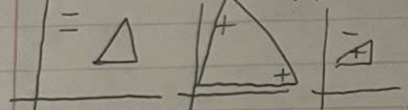
Problem #1b ~~#2~~:

We conside H with triangular decision boundary of any shape. Points inside classified pos, outside neg. We again graphically represent this problem in 2D.



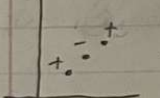
It is trivial to see that with a single datapoint, it can be perfectly classified no matter the location or label. $VCD(H) \geq 1$

The same applies to 2 data points:



Again, trivial to see that any label combination can be represented anywhere in the problem space. $VCD(H) \geq 2$

We then consider a case of 3 datapoints where each sample lies along the same line in space.

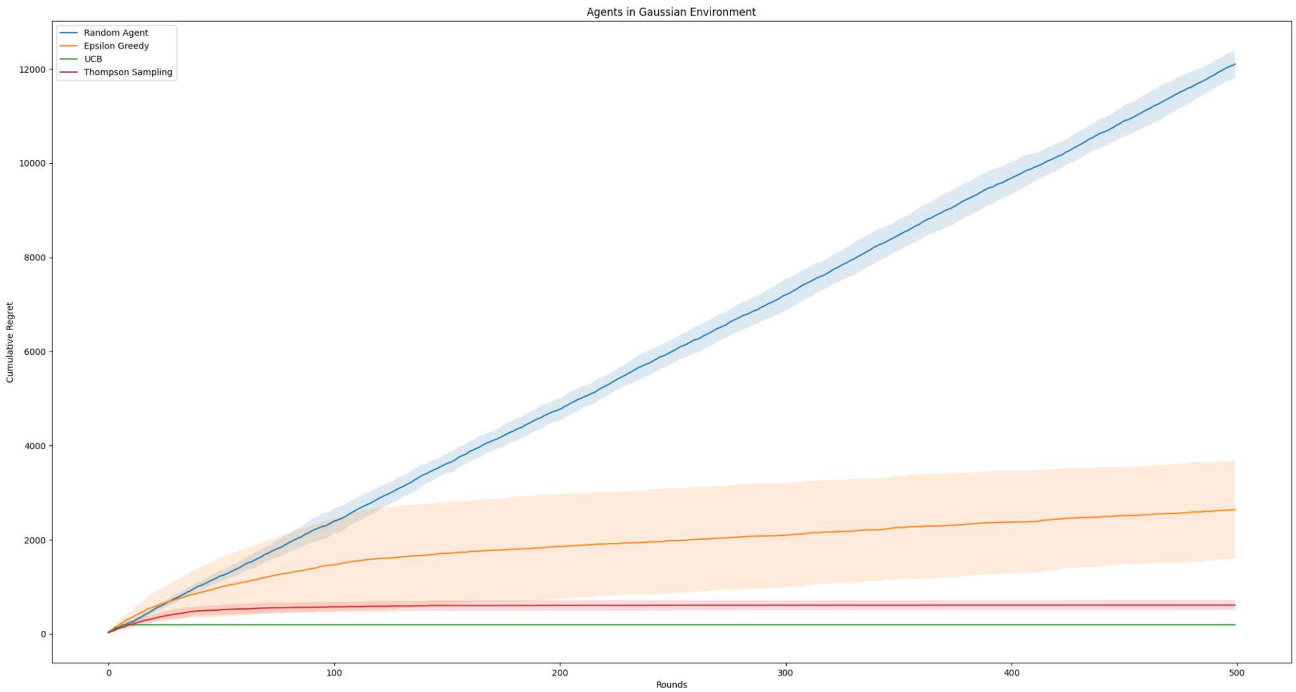


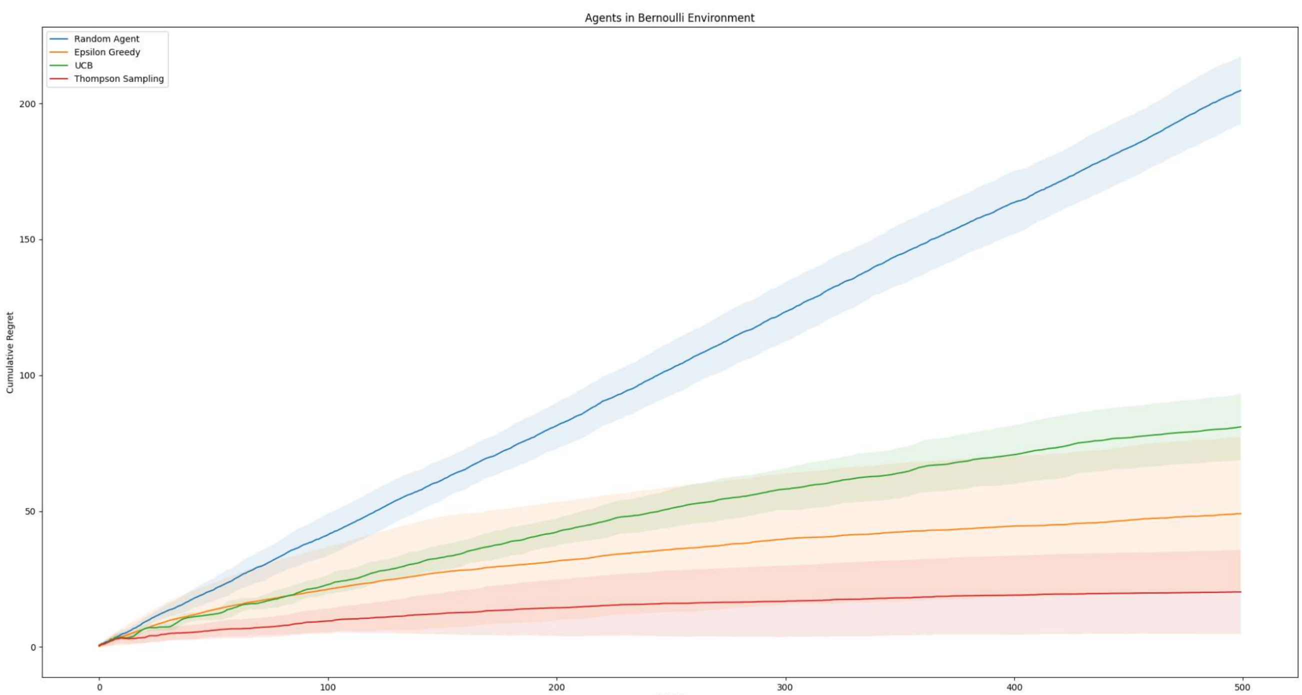
There is no possible triangle that can satisfy the labelling to the left. A triangle with any volume will misclassify the middle point necessarily and even a triangle with no volume would be the exact line each point lies along, misclassifying the middle point again. Thus, $VCD(H) < 3$

Thus, $VCD(H) = 2$

Q.E.D.

Problem #2:





The above two plots show the average cumulative regret over 500 rounds for each of the algorithms for both gaussian and bernoulli distribution environments with the error bars being the standard deviation. Looking at the gaussian environment, we can see that the bounds are pretty similar to what was talked about in class. All of the 3 algorithms are sub-linear and random is linear as it should be. UCB just kind of hard cuts off which also follows because of the “upper confidence bound”. Epsilon greedy having a bit less of a cutoff than the other two also tracks with what was mentioned in class. The bernoulli environment follows almost the same pattern. Epsilon greedy looks pretty similar to the gaussian environment, and thompson sampling takes a little bit longer to converge. The biggest difference is UCB which has a worse curve than epsilon greedy. It's possible that this distribution is just harder to find the upper confidence bound than the gaussian distribution. Additionally, in the bernoulli environment, all of the algorithms have much higher standard deviations than in the gaussian environment, showing that they are not converging as well. The random agent is linear as it should be. From the two plots, we can conclude that how well these algorithms work depend on the underlying distribution. While they may work very well on a Gaussian distribution, they may not work as well on another distribution such as bernoulli.