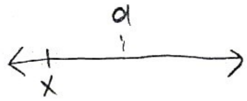
Gurkan Kilicaslan 5810725 kilia064@umn.edu

Problem 1

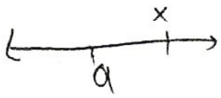
a)

x with label 0



we always have an "a" to the right to classify x as 0.

x with label 1



we always have an "a" to the left to classify x as 1.

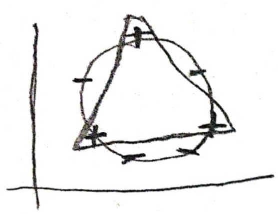
$\Rightarrow VCdim(H) \geq 1$ ①

Consider x_1 with label 1 and x_2 with label 0 and place them as follows

there is no "a" that correctly classifies x_1 and x_2 because by definition, points to the right of "a" are labeled as 1 and 0 otherwise. However, in the above case there is no "a" that satisfies this. $\Rightarrow VCdim(H) < 2$ ②

$\longrightarrow \boxed{VCdim(H) = 1}$

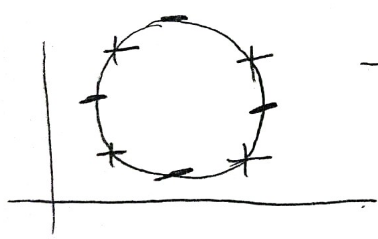
b)



Up to 7 points can be seperated by a triangle.

However, 8 or more points cannot be seperated by any triangle because order of the points matter a lot

For example:



$\rightarrow$ Points cannot be seperated by any triangle.

The reason for 7 datapoints, is that maximum numbers of groups of points can be 4 and 3. Class with 3 poins can be seperated with 3 edges of the triangle. $\boxed{VCdim(H) = 7}$

Problem 2

a)

$$E_{x,y,D}\left[\left(\overset{(1)}{f_D(x)} - \overset{(2)}{\bar{f}(x)}\right)\left(\overset{(a)}{\bar{f}(x)} - \overset{(b)}{y}\right)\right] =$$

$$= \underset{(1)(a)}{E\left[f_D(x)\,\bar{f}(x)\right]} - \underset{(2)(a)}{E\left[\bar{f}(x)^2\right]} - \underset{(1)(b)}{E\left[f_D(x)\,y\right]} + \underset{(2)(b)}{E\left[\bar{f}(x)\,y\right]}$$

$$= \bar{f}(x)\,E\left[f_D(x)\right] - \bar{f}(x)^2 - E\left[f_D(x)\,y\right] + E\left[E\left[f_D(x)\right]y\right]$$

$$= \cancel{\bar{f}(x)^2} - \cancel{\bar{f}(x)^2} - \cancel{E\left[f_D(x)\,y\right]} + \cancel{E\left[f_D(x)\,y\right]} = \underline{\underline{0}}$$

b)

$$E_{x,y}\left[\left(\overset{(1)}{\bar{f}(x)} - \overset{(2)}{\bar{y}(x)}\right)\left(\overset{(a)}{\bar{y}(x)} - \overset{(b)}{y}\right)\right] = \underset{(1)(a)}{E\left[\bar{f}(x)\,\bar{y}(x)\right]} - \underset{(2)(a)}{E\left[\bar{f}(x)\,y\right]} - \underset{(1)(b)}{E\left[\bar{y}(x)^2\right]} + \underset{(2)(b)}{E\left[\bar{y}(x)\,y\right]}$$

$$= \bar{f}(x)\,\bar{y}(x) - \bar{f}(x)\,E[y] - \bar{y}(x)^2 + \bar{y}(x)\,E[y]$$

$$= \bar{f}(x)\,\bar{y}(x) - \bar{y}(x)^2 + \left(\bar{y}(x) - \bar{f}(x)\right)E(y)$$

$$= \cancel{\bar{f}(x)\,\bar{y}(x)} - \cancel{\bar{y}(x)^2} + \cancel{\bar{y}(x)^2} - \cancel{\bar{f}(x)\,\bar{y}(x)}$$

$$= \underline{\underline{0}}$$

Problem 3
a)

$$\frac{\delta}{2} = exp\Big(-\frac{t^2/2}{\sum_{i=1}^{n} E[X_i^2] + Mt/3}\Big)$$
$$ln\big(\frac{\delta}{2}\big) = -\frac{t^2/2}{\sum_{i=1}^{n} E[X_i^2] + Mt/3}$$
$$ln\big(\frac{2}{\delta}\big) = \frac{t^2/2}{\sum_{i=1}^{n} E[X_i^2] + Mt/3}$$
$$ln\big(\frac{2}{\delta}\big)\big(\sum_{i=1}^{n} E[X_i^2] + Mt/3\big) = \frac{t^2}{2}$$
$$ln\big(\frac{2}{\delta}\big)\big(\sum_{i=1}^{n} E[X_i^2] + Mt/3\big) - \frac{t^2}{2} = 0$$
$$\frac{t^2}{2} - ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3}t\big) - ln\big(\frac{2}{\delta}\big)\big(\sum_{i=1}^{n} E[X_i^2]\big) = 0$$

Last equation is a second degree quadratic equation and the solution to the equation is the following:

$$t = ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3}\big) + \Big(\sqrt{ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3}\big)}\Big)^2 + \sqrt{2ln\big(\frac{2}{\delta}\big)\sum_{i=1}^{n} E[X_i^2]}$$
$$t = 2ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3}\big) + \sqrt{2ln\big(\frac{2}{\delta}\big)\sum_{i=1}^{n} E[X_i^2]}$$

We are given that

$$\mathbb{P}\Big[\big|\sum_{i=1}^{n} X_i > t\big|\Big] = \delta$$
$$\mathbb{P}\Big[\big|\sum_{i=1}^{n} X_i \leq t\big|\Big] = 1-\delta$$
$$\big|\sum_{i=1}^{n} X_i\big| \leq t$$
$$\big|X_i\big| \leq \frac{t}{n}$$

Substitute t:

$$\big|X_i\big| \leq \frac{2ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3}\big) + \sqrt{2ln\big(\frac{2}{\delta}\big)\sum_{i=1}^{n} E[X_i^2]}}{n}$$
$$\big|X_i\big| \leq 2ln\big(\frac{2}{\delta}\big)\big(\frac{M}{3n}\big) + \sqrt{\frac{2ln\big(\frac{2}{\delta}\big)\sum_{i=1}^{n} E[X_i^2]}{n}}$$
$$\big|X_i\big| \leq \sqrt{\frac{2\sum_{i=1}^{n} E[X_i^2]log\big(\frac{2}{\delta}\big)}{n}} + \frac{2Mlog\big(\frac{2}{\delta}\big)}{3n}$$

b)

Normalize X_i with the mean 0:

$$X_i = \mathbf{1}\{\hat{y}_i \neq y_i\} - E[\mathbf{1}\{\hat{y}_i \neq y_i\}],$$

$$E[X_i] \leq E[(\mathbf{1}\{\hat{y}_i \neq y_i\})^2] \leq E[\mathbf{1}\{\hat{y}_i \neq y_i\}] = R(h), \qquad M \leq 1.$$

Using the equation (3) from the homework file and replacing δ with $\delta/|\mathcal{H}|$(union bound):

$$|\hat{R}(h) - R(h)| \leq \sqrt{\frac{2R(h)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{2log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

$\hat{h}$ is the minimizer of $\hat{R}(h)$, so: $\hat{R}(\hat{h}) - \hat{R}(h^*) \leq 0$:

$$\hat{R}(\hat{h}) - \hat{R}(h^*) \leq \hat{R}(h^*) - R(h^*) - \hat{R}(\hat{h}) + R(\hat{h}) \leq |\hat{R}(h^*) - R(h^*)| + |-\hat{R}(\hat{h}) + R(\hat{h})|$$

$$\leq \sqrt{\frac{2R(\hat{h})log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{2log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n} + \frac{2log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

$$(*) \qquad \sqrt{\frac{2(R(\hat{h}) - R(h^*))log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + 2\sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{4log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

We know that if $a, b \geq 0$, then $1/2(a+b) \geq \sqrt{ab}$. Applying this relation to the first term in the above equation where $a = R(\hat{h}) - R(h^*)$ and $b = \frac{2log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}$:

$$\sqrt{\frac{2(R(\hat{h}) - R(h^*))log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} \leq \frac{1}{2}(R(\hat{h}) - R(h^*)) + \frac{log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}$$

Substituting this back in (*):

$$\sqrt{\frac{2(R(\hat{h}) - R(h^*))log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + 2\sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{4log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

$$\leq \frac{1}{2}(R(\hat{h}) - R(h^*)) + 2\sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{7log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

All in all we have:

$$R(\hat{h}) - R(h^*) \leq \frac{1}{2}(R(\hat{h}) - R(h^*)) + 2\sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{7log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

$$\implies \frac{1}{2}(R(\hat{h}) - R(h^*)) \leq 2\sqrt{\frac{2R(h^*)log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{n}} + \frac{7log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{3n}$$

$$\implies R(h^*) - R(\hat{h}_n) \leq c_1\sqrt{\frac{R(h^*)log\left(\frac{|\mathcal{H}|}{\delta}\right)}{n}} + c_2\frac{log\left(\frac{|\mathcal{H}|}{\delta}\right)}{n}$$

Problem 4

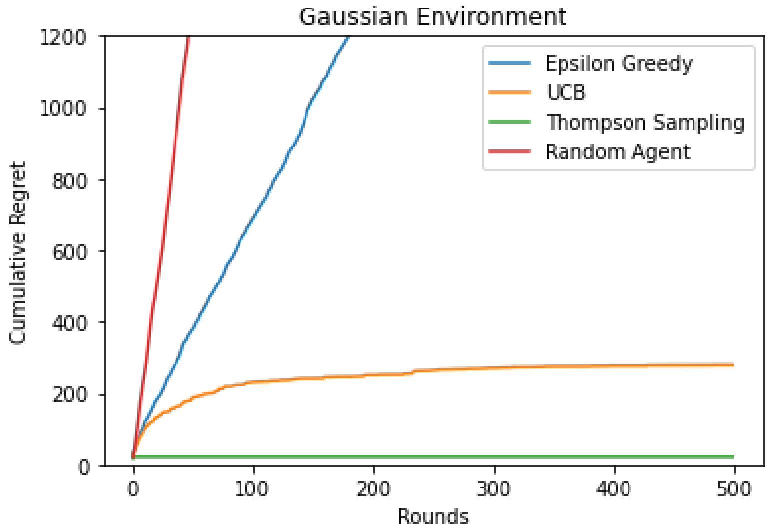


FIGURE 1. Average cumulative regrets by rounds under Gaussian Environment

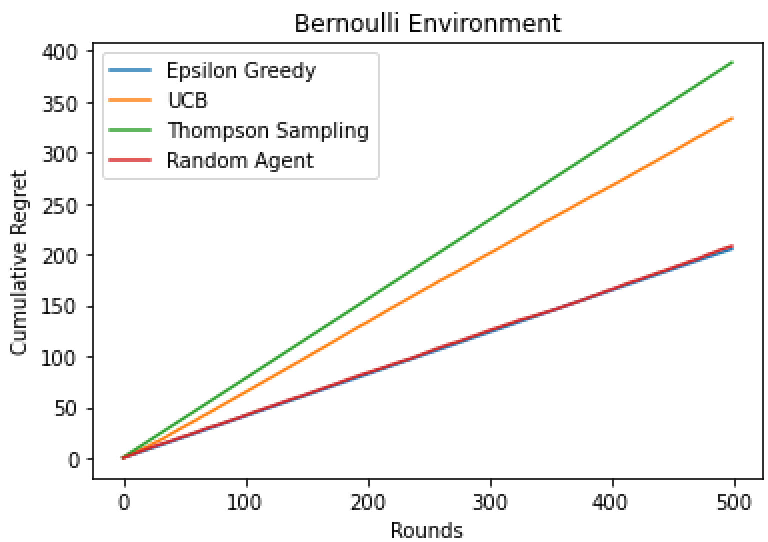


FIGURE 2. Average cumulative regrets by rounds under Bernoulli Environment

Out of the four models, Thompson Sampling displayed the least amount of regret under Gaussian Environment, whereas Random Agent had the highest and most linear regret. Epsilon Greedy had the second-largest regret.

In the Bernoulli Environment, Epsilon Greedy demonstrated the least amount of regret despite being extremely similar to Random Agent, UCB had the second-biggest regret, and Thompson Sampling had the largest regret that was linear.

As it should be, Random Agent is always linear. None of the agents in the Bernoulli Environment are sub-linear; rather, they are all linear. Remarkably, in contrast to the other two, Random Agent and Epsilon Greedy have far lower rates of regret accumulation. This results shows they are doing better when it comes to Bernoulli Environment. I wouldn't say the algorithms have shown great success especially in Bernoulli Environment.