

# CS 260B Homework 2

Hanna Co

Collaborators: Isha Gonugunta

Due: April 27, 2022

# 1 Problem 1

We start with  $S_* = 1, 2, \dots, d$

For each example  $(x_i, y_i)$ :

if  $y_i = 1$  and  $AND(x_i) \neq y_i$ , then we remove  $j$  from  $S_*$  where  $x_j = 0$

The time complexity for each example is  $O(d)$ , for iterating through each element of  $x$ .

This mistake bound is  $d$ , because since the dimension on  $S_*$  is  $d$ , then we can only make at most  $d$  mistakes.

## 2 Problem 2

Let's say we have two experts, X and Y, where X always predicts *up* and Y always predicts *down*. Our algorithm is deterministic, so we can make it so that our algorithm is always wrong. Thus, since we only have two experts who never make the same prediction, at least one of them has to be correct at least half of the time. That is,  $L \leq \frac{T}{2}$ . Our algorithm is always incorrect, so  $L_{our} = T$ . Thus,  $L_{best} \leq \frac{T}{2}$  and  $L_{our} = T$ , then  $L_{our} \geq 2 * L_{best}$ .

### 3 Problem 3

We can take our multiplicative weights algorithm, and adjust the way we update the weights:

$$\text{if correct, } w(t, j) = w(t, j - 1)$$

$$\text{if wrong, } w(t, j) = w(t - 1, j) * (1 - \epsilon)^{L(t-1, j)}$$

where  $L(t - 1, j) = |\text{prediction} - \text{actual}|$

For some expert  $i$ :

$$w(0) = 1$$

$$w(1) = w(0)(1 - \epsilon)^{L(0)} = (1 - \epsilon)^{L(0)}$$

$$w(2) = w(1)(1 - \epsilon)^{L(1)}$$

We substitute in  $w(1)$  to get:

$$w(2) = (1 - \epsilon)^{L(0)}(1 - \epsilon)^{L(1)}$$

which simplifies to:

$$w(2) = (1 - \epsilon)^{L(0)+L(1)}$$

Thus, the total weight for expert  $i$  after  $T$  days is

$$w(T, i) = (1 - \epsilon)^{\sum_{t=1}^T L(t, i)}$$

We can write  $W(t)$  as the summation of all the weights on a particular day for all the experts:

$$W(t) = \sum_{i=1}^d w(t, i)$$

$$W(t) = \sum_{i=1}^d w(t - 1, i)(1 - \epsilon)^{L(t, i)}$$

We know that

$$(1 - \epsilon)^x \leq (1 - \epsilon x)$$

Then,

$$\begin{aligned} \sum_{i=1}^d w(t-1, i)(1 - \epsilon)^{L(t, i)} &\leq \sum_{i=1}^d w(t-1, i)(1 - \epsilon L(t, i)) \\ &= \sum_{i=1}^d w(t-1, i) - \epsilon \sum_{i=1}^d w(t-1, i)L(t, i) \\ &= W(t-1) - \epsilon \sum_{i=1}^d w(t-1, i)L(t, i) \end{aligned}$$

From lecture, we have that  $L(t) = \frac{1}{W(t-1)} \sum_{i=1}^d w(t-1, i)L(t, i)$ . Thus, we have

$$W(t) \leq W(t-1) - \epsilon L(t)W(t-1)$$

$$W(t) \leq W(t-1)(1 - \epsilon L(t))$$

Knowing that  $1 - x \leq e^{-x}$ , then we can say

$$W(t) \leq W(t-1)e^{-\epsilon L(t)}$$

Therefore,

$$W(T) \leq e^{-\epsilon A(T)}$$

where  $A(T)$  is our total expected loss. Also from lecture, we have that  $W(T) \geq (1 - \epsilon)^{A_*(T)}$  where  $A_*(T)$  is the loss of the best expert, giving us

$$(1 - \epsilon)^{A_*(T)} \leq d * e^{-\epsilon A(T)}$$

$$A_*(T) * \ln(1 - \epsilon) \leq \ln(d) - \epsilon A(T)$$

$$(\epsilon)(A(T)) \leq (-\ln(1 - \epsilon)) * A_*(T) + \ln(d)$$

$$A(T) \leq \left(\frac{-\ln(1 - \epsilon)}{\epsilon}\right)A_*(T) + \frac{\ln(d)}{\epsilon}$$

Therefore, since we have that  $\frac{-\ln(1-x)}{x} \leq 1+x$ ,

$$A(T) \leq (1+\epsilon)A_*(T) + \frac{\ln(d)}{\epsilon}$$

## 4 Problem 4

We can approach this problem as a learning with experts problem, with each component of  $x$  as an expert's guess, and the components of  $w_*$  as weights. We can model the multiplicative weights algorithm, where if an expert is wrong, we decrease their weight, and if they are correct, we increase their weight. The new algorithm is as follows:

For every example  $(x_t, y_t)$ :

Pick an "expert" (component of  $x_t$ ) with probability proportional to its "weight" (component of  $w_*(t)$ )

Follow this "expert"

We update the "weights" for each expert  $i$ :

$$w_*(t+1, i) = w(t, i)(1 - \epsilon L(t, i)), \text{ where } L(t, i) = -y(t, i) * x_t(i)$$

From lecture, we have that

$$0 \leq \mathbb{E}[L(t)] \leq L_*(T) + 2\sqrt{T * \ln(d)}$$

where  $L_*(T)$  is the total loss of the best expert.

We also know that

$$\mathbb{E}[\text{loss on day } t] = \sum_{i=1}^d \Pr[\text{picking expert } i] * L(t, i)$$

$\Pr[\text{picking expert } i]$  is proportional to weight:  $\frac{w_{t-1}(i)}{\sum_{j=1}^d w_{t-1}(j)}$ , but we know that the sum of the entries of  $w_{t-1}$  add up to 1, thus

$$\Pr[\text{picking expert } i] = w_{t-1}(i)$$

So we have

$$\mathbb{E}[\text{loss on day } t] = \sum_{i=1}^d w_{t-1}(i) * L(t, i)$$

but  $L(t, i) = -y(t, i) * x_t(i)$ , giving us

$$\mathbb{E}[\text{loss on day } t] = \sum_{i=1}^d w_{t-1}(i) * -y(t, i) * x_t(i)$$

This, we can rewrite as

$$\mathbb{E}[\text{loss on day } t] = -y(t) * \langle w_{t-1}, x_t \rangle$$

We can say that

$$\begin{aligned} 0 &\leq \mathbb{E}[L(t)] + 2\sqrt{T * \ln(d)} \\ 0 &\leq \sum_{t=0}^T \mathbb{E}[\text{loss on day } t] + 2\sqrt{T * \ln(d)} \\ 0 &\leq \sum_{t=0}^T -y(t, i) * \langle w_{t-1}, x_t \rangle + 2\sqrt{T * \ln(d)} \\ 0 &\leq M * -y(i) * \langle w_{t-1}, x_t \rangle + 2\sqrt{T * \ln(d)} \end{aligned}$$

Where M is the number of mistakes we make. This inequality becomes

$$\begin{aligned} 0 &\leq M * -\gamma + 2\sqrt{T * \ln(d)} \\ M * \gamma &\leq 2\sqrt{T * \ln(d)} \\ M &\leq \frac{2\sqrt{T * \ln(d)}}{\gamma} \end{aligned}$$