

## Lecture 19: Multiplicative Weight Update and LP

Instructor: *Alex Andoni*Scribes: *Garrison Grogan, Shunhua Jiang*

## 1 Problem Setting

Think about any prediction problem, for example, predicting the stock price. We want to rely on the advices from a group of experts. We define the following notations:

- $n$  experts,
- $t \in \{1, \dots, T\}$ : time.
- for  $i \in [n]$ ,  $t \in [T]$ ,  $f_i^t := \begin{cases} 1, & \text{expert } \#i \text{ is wrong on day } t, \\ 0, & \text{otherwise.} \end{cases}$
- $m_i^t :=$  number of errors that expert  $\# i$  made up to day  $t$ ,
- $M^t :=$  number of errors we make.

Our goal: do as close as possible to the best expert, where the best expert is defined as

$$\arg \min_i m_i^T.$$

## 2 Examples of Bad Algorithms

**(1) Majority vote:** Follow the majority of the advices by the experts.

Consider the following example: There are in total 3 experts, and the 1st and 2nd experts always give the advice to sell, while the 3rd expert always gives the advice to buy. And the right strategy is to buy everyday. In this example the majority vote algorithm is always wrong, and its mistakes are  $M^t$ .

**(2) Follow the expert that was right yesterday:** Consider the following example. There are two experts. The advices by the experts and the right strategy are as follows (where  $B$  means buy and  $S$  means sell):

- 1st:  $S, B, S, B, \dots$ ;
- 2nd:  $B, S, B, S, \dots$ ;
- right:  $S, S, S, S, \dots$ .

Observe that except the first day, this algorithm is always wrong by following the expert that was right yesterday, since he is no longer right today. Thus the total number of mistakes is  $M^T = T$ .

### 3 Weighted Majority Algorithm

Intuitively, this algorithm is a combination of the previous two algorithms: it follows the majority vote of the experts that were more right. The algorithm assigns weights to experts, and  $w_i^t$  denotes the weight for expert  $i$  at time  $t$ .

---

#### Algorithm 1 Weighted Majority Algorithm

---

```

1: Start:  $w_i^0 = 1, \forall i \in [n]$ .
2: for  $t = 1, 2, \dots, T$  do
3:   Update  $w_i^{t+1} = w_i^t(1 - \epsilon \cdot f_i^t)$ , where  $\epsilon < 1$  is a parameter TBD.
4:   Weighted majority:  $\sigma_B^t = \sum_{i: \text{expert } i \text{ says } B} w_i^t, \quad \sigma_S^t = \sum_{i: \text{expert } i \text{ says } S} w_i^t$ .
5:   if  $\sigma_B^t > \sigma_S^t$  then
6:     Decision at time  $t$ : buy
7:   else
8:     Decision at time  $t$ : sell
9:   end if
10: end for

```

---

#### Analysis.

**Theorem 1.** *We have the following upper bound on the final number of mistakes.*

$$M^T \leq 2(1 + \epsilon)m_i^T + \frac{2 \ln n}{\epsilon}, \quad \forall i.$$

And this implies that  $M^T \leq 2(1 + \epsilon) \min_{i \in [n]} m_i^T + \frac{2 \ln n}{\epsilon}$ .

A remark about why we need to have the additive error  $\frac{2 \ln n}{\epsilon}$ : Consider the example where we have  $n - 1$  random experts and 1 expert that is always right. In the first few days, we do not have enough information to know which expert is the right one. Indeed, on the  $t$ -th day, in expectation there are  $2^{-t}$  experts that were always right in previous days. So only until the  $\ln n$  day we have gathered enough information to find the right expert.

*Proof.* We prove the theorem by giving a lower bound and an upper bound of the following potential function  $\Phi$ .

$$\forall t \in [T], \quad \Phi_t := \sum_{i=1}^n w_i^t; \quad \Phi_0 := n.$$

(1)  $\Phi_{t+1} \geq w_i^{t+1} = (1 - \epsilon)^{m_i^t}, \quad \forall i.$

The first step follows from the definition of  $\Phi_{t+1}$ , and the second step follows from  $w_i^0 = 1$ , and the fact that the weight of  $i$  is multiplied by  $(1 - \epsilon)$  for  $m_i^t = (\# \text{ of errors})$  times.

(2)  $\Phi_{t+1} \leq n \cdot (1 - \frac{\epsilon}{2})^{M^t}.$

If at time  $t$  we don't make an error:  $\Phi_{t+1} \leq \Phi_t.$

If at time  $t$  we made an error: weight of wrong experts  $\geq$  weight of right experts  $\geq \frac{1}{2}\Phi_t.$

Combining these two cases, we have

$$\begin{aligned}
\Phi_{t+1} &= \sum_{i=1}^n w_i^{t+1} \\
&= \sum_{i:\text{right}} w_i^t + \sum_{i:\text{wrong}} w_i^t(1 - \epsilon) \\
&= \sum_{i=1}^n w_i^t - \epsilon \cdot \sum_{i:\text{wrong}} w_i^t \\
&\leq \Phi_t - \epsilon \cdot \frac{1}{2} \Phi_t = \Phi_t \left(1 - \frac{\epsilon}{2}\right).
\end{aligned}$$

Thus we have  $\Phi_{t+1} \leq \Phi_0 \cdot \left(1 - \frac{\epsilon}{2}\right)^{M^t} = n \cdot \left(1 - \frac{\epsilon}{2}\right)^{M^t}$ .

Putting (1) and (2) together, we have

$$(1 - \epsilon)^{m_i^t} \leq \Phi_{t+1} \leq n \cdot \left(1 - \frac{\epsilon}{2}\right)^{M^t}. \quad (1)$$

Using  $1 - x \leq e^{-x}$ , we have  $n \cdot \left(1 - \frac{\epsilon}{2}\right)^{M^t} \leq n \cdot e^{-\frac{\epsilon}{2} \cdot M^t}$ . And using  $\ln(1 - \epsilon) \geq -\epsilon - \epsilon^2$ , we have  $-\epsilon(1 + \epsilon)m_i^t \leq m_i^t \cdot \ln(1 - \epsilon)$ . Plugging these two inequality to Eq. 1, we have

$$-\epsilon(1 + \epsilon)m_i^t \leq \ln n - \frac{\epsilon}{2} \cdot M^t,$$

and this implies the desired bound

$$M^t \leq 2(1 + \epsilon)m_i^t + \frac{2 \ln n}{\epsilon}.$$

□

The Weighted Majority Algorithm is good but we can do better. Our next algorithm will remove this factor of 2.

## 4 Multiplicative Weights Update

The algorithms is the same as before but the decision is as follows:

Define  $p_i^t = \frac{w_i^t}{\sum_{j=1}^n w_j^t}$ , the probability of following expert  $i$  at time step  $t$ .

Next choose  $i$  from the distribution  $(p_1^t, \dots, p_n^t)$  and follow it's decision. Now the number of errors we make,  $M^T$  is a random variable.

**Analysis.** Each  $m_i^t$  is deterministic.  $M^T \triangleq$  expected # of errors we make  $= \sum_{t=1}^T \sum_{i=1}^n p_i^t f_i^t = \sum_{t=1}^T \langle p^t, f^t \rangle$

We can generalize  $f_i^t \in [-1, +1]$ , and  $m_i^t = \sum_{j=1}^t f_i^j$

**Theorem 2.**  $M^T \leq m_i^t + \frac{\ln n}{\epsilon} + \epsilon T$ .

Note the factor of 2 is no gone!

*Proof.*  $\Phi^{T+1} \triangleq \sum_{i=1}^n w_i^{T+1}$

Now note the following:

$$1. \Phi^{T+1} \geq w_i^{T+1} \geq 1 * \prod_{t=1}^T (1 - \epsilon f_i^t).$$

The  $(1 - \epsilon f_i^t)$  term is how much we scale down weight  $w_i^t$  at time  $t$

$$2. \Phi^{T+1} \leq n e^{-\epsilon M^T}$$

Now consider the following calculation:

$$\begin{aligned} p_i^t = \frac{w_i^t}{\Phi^t} &\implies \Phi^{T+1} = \sum_{i=1}^n w_i^{T+1} = \sum_{i=1}^n w_i^T * (1 - \epsilon f_i^T) = \sum_{i=1}^n w_i^T - \epsilon \sum_{i=1}^n \frac{w_i^T f_i^T}{\Phi^T} * \Phi^T \\ &= \Phi^T - \epsilon \Phi^T \langle p^T, f^T \rangle \\ &= \Phi^T (1 - \epsilon \langle p^T, f^T \rangle) \end{aligned}$$

$$\implies \Phi^{T+1} \leq \Phi^T e^{-\epsilon \langle p^T, f^T \rangle}, \text{ which is}$$

$$\begin{aligned} &\leq \Phi^0 e^{-\epsilon \sum_{t=1}^T \langle p^t, f^t \rangle} \\ &= \Phi^0 e^{-\epsilon M^T} = n e^{-\epsilon M^T} \end{aligned}$$

Now using 1 and 2, we get  $\Phi^{T+1} \leq n * e^{-\epsilon M^T}$  and  $\prod_{t=1}^T (1 - \epsilon f_i^t) \leq \Phi^{T+1} \leq n e^{-\epsilon M^T}$

Next we see that  $\ln n - \epsilon M^T \leq \sum_{t=1}^T \ln(1 - \epsilon f_i^t) \geq -\sum_{t=1}^T [\epsilon f_i^t + \epsilon^2 (f_i^t)^2]$

Finally we see that  $M^T \leq \frac{\ln n}{\epsilon} + \sum_{t=1}^T [f_i^t + \epsilon] \leq m_i^T + \frac{\ln n}{\epsilon} + \epsilon T$  for the optimal expert  $i$ . □