COMS 4995-8: Advanced Algorithms (Spring'21) Apr 8, 2021

Lecture 24: Multiplicative Weights Update

Instructor: *Alex Andoni* Scribes: *Arkadiy Saakyan as5423, Andrew Schreiber ajs2409*

1 Setting

Imagine the following setting: There are n experts advising to buy or sell a stock at time t. Let $f_i^t = 1$ if the expert i is wrong at time t, and 0 otherwise. Let T be the total number of observations.

Definition 1. *Goodness of expert:*
Let m_i^t *= number of errors by expert* i *at time* t*.*
Let M^t *= total number of errors at time* t*.*
Best expert $i = \arg\min_{i\in[n]} m_i^T$

Goal: bound M^T as a function of best expert prediction. Ideally, $M^T \leq m_{best}^T$.

2 Examples of bad algorithms

1. Choosing majority. Suppose the following scenario, $n = 3$: Here we incur T errors even though there

expert 1	B	B	B
expert 2	B	B	B
expert 3	S	S	S
true	S	S	S

is an expert that is always correct!
2. Following yesterday. Suppose the following scenario, $n = 2$: Since the yesterday's correct expert is

expert 1	B	S	B	S
expert 2	S	B	S	B
true	B	B	B	B

always wrong on the current day, we end up being wrong every day!

3 Weighted Majority Algorithm

Algorithm 1: Weighted Majority Algorithm

> initialize $w_i^0 = 1$;
> **for** $t = 1 \dots T$ **do**
> $\sigma_B^t = \sum_{i:B} w_i^t$;
> $\sigma_S^t = \sum_{i:S} w_i^t$;
>   **if** $\sigma_B^t > \sigma_S^t$ **then**
>   buy at t;
>   **else**
>   sell at t;
>   **end**
>   **for** $i = 1 \dots n$ **do**
> $f_i^t = 1$ if i is wrong;
> $w_i^{t+1} = w_i^t(1 - \epsilon \cdot f_i^t)$
>   **end**
> **end**

Theorem 2. $\forall i \in [n] : M^T \leq 2(1+\epsilon) \cdot m_i^t + \frac{2 \ln n}{\epsilon}$

Proof. We will use the idea of a "potential".

Let $\Phi^t = \sum_{i \in [n]} w_i^t$

Want: $m_i^T \leq \Phi^{T+1} \leq M^T$

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

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

   To see that 2 is true, suppose we make an error at time t.
   Then $\sigma_{corr} \leq \sigma_{wrong}$
   Then $\sum_{i:corr} w_i^t \leq \sum_{i:wrong} w_i^t$
   Then we have that

$$
\begin{aligned}
\Phi^{t+1} &= \sum_{i:corr} w_i^t + \sum_{i:wrong} w_i^t(1-\epsilon) \\
&\leq \Phi^T - \epsilon \sum_{i:wrong} w_i^t \\
&\leq \Phi^t - \frac{\epsilon}{2} \cdot \Phi^t \\
&= \Phi^t(1 - \frac{\epsilon}{2})
\end{aligned}
$$

By 1 and 2, we have that

$$(1-\epsilon)^{m_i^T} \leq \Phi^{T+1} \leq n(1 - \frac{\epsilon}{n})^{M^T}$$

We know that $n(1-\frac{\epsilon}{n})^{M^T} \leq n \cdot e^{-\frac{\epsilon}{2}M^T}$
We also know that $\ln 1-\epsilon \geq -\epsilon-\epsilon^2$
Therefore,

$$m_i^T(-\epsilon-\epsilon^2) \leq \ln n - \frac{\epsilon}{2}M^T$$

Hence, we have that

$$\frac{\epsilon}{2}M^T \leq \epsilon(1+\epsilon)\cdot m_i^T + \ln n$$
$$\implies M^T \leq 2(1+\epsilon)\cdot m_i^T + \frac{2\ln n}{\epsilon}$$

□

4 MWU: Multiplicative Weights Update Algorithm

In the Weighted Majority Algorithm we used f_i^t to be a binary flag indicating if expert i was wrong at time t (1 indicating incorrect). Let us now redefine $f_i^t \in [-1,+1]$ s.t. f measures to what degree a particular expert was right or wrong (-1 being the most correct, 1 being the most incorrect). Let:

$$m_i^t = \sum_{j\leq t} f_i^j$$

Multiplicative Weights Update Algorithm: Same as Weighted Majority Algorithm except choose to follow expert i randomly, proportional to its weight. More Formally:

$$p_i = \frac{w_i^t}{\Phi^t}, \quad p^t = (p_1^t, p_2^t, \ldots, p_n^t)$$

$$w_i^{t+1} = w_i^t(1-\epsilon f_i^t)$$

Its worth explicitly noting that the fact that f can give negative values means when experts get things right we increase their weights, as opposed to the old algorithm where we could only decrease or keep the weights the same.

Theorem 3. *$M^T \triangleq \mathbb{E}[\#$ errors we make$] = \sum_{t=1}^T \mathbb{E}[f_i^t] = \sum_{t=1}^T \sum_{i=1}^n p_i^t f_i^t = \sum_t \langle p^t, f^t\rangle$*
Then: $M^T \leq m_i^t + \epsilon T + \frac{ln(n)}{\epsilon}$.

Proof. Let: $\Phi^t = \sum_i w_i^t$

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

2. $\Phi^1 = n$

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

Thus we have: $\Phi^{T+1} = \Phi^1 \cdot \prod_{t=1}^{T}(1 - \epsilon \cdot \langle p^t, f^t \rangle)$. Thus:

$$\Phi^{T+1} \leq n \prod_{t=1}^{T} e^{-\epsilon \cdot \langle p^t, f^t \rangle} = n \cdot e^{-\epsilon \sum_t \langle p^t, f^t \rangle} = n \cdot e^{-\epsilon M^T}$$

So combining 1 and 2 we have:

$$\prod_t (1 - \epsilon f_i^t) \leq n \cdot e^{-\epsilon M^T}$$

$$\sum_t \ln(1 - \epsilon f_i^t) \leq \ln(n) - \epsilon M^T$$

And now trying to lower bound the left side we have:

$$\sum_t \ln(1 - \epsilon f_i^t) \geq \sum_t -\epsilon f_i^t - \epsilon^2 (f_i^t)^2 \geq \sum_t -\epsilon f_i^t - \epsilon^2 \geq -\epsilon \cdot m_i^t - \epsilon^2 \cdot T$$

Thus combining the previous equations we have:

$$-\epsilon \cdot m_i^t - \epsilon^2 \cdot \leq \ln(n) - \epsilon M^T$$

$$M^T \leq m_i^T + \epsilon \cdot T + \frac{\ln(n)}{\epsilon}$$

□

Since this is true $\forall i \in [n] \implies M^T \leq \min_i m_i^T + \epsilon \cdot T + \frac{\ln(n)}{\epsilon}$

<u>Remark</u>: if $f_i^t \geq 0 \implies M^T \leq (1+\epsilon) m_i^T + \frac{\ln(n)}{\epsilon}$