Today

Multiplicative Weight Update Method (part I)

CS5.205.1
Toolkit in TCS
- Lecture #15
(7 Apr '21)
Instructor: Prahladh Harsha

Online Algorithms

Multiplicative Weight Update Method.

Core of several algorithms

- Machine Learning (Adaboost, Winnow)
- Optimization (solve linear programming)
- Algorithms (faster algs for LPs & SDPs)
- Complexity (Impagliazzo's proof of Hardcore Set Lemma)
- Game Theory (computing zero-sum games)

References: (1) Arora - Hazan - Kale
(2) Gupta - O'Donnell (CMU)
(3) Vazirani - Rao (Berkeley)

Warmup Case : Infallible Expert

n experts → predictions (Boolean valued)
0/1
(up/down, win/loss)

Minimize # losses.

1. Initialize $\mathcal{E} = \{E_1, \ldots, E_n\}$
   (set of experts who haven't committed error)
2. For $t \leftarrow 1$ to T.
   (a). Act according to the majority predictions of experts in $\mathcal{E}$.
   (b). Observe the outcome
   (c) Eliminate from $\mathcal{E}$ all experts w/ an incorrect prediction

Analysis: (*) Each time, I make a mistake $\frac{1}{2}$ the experts in $\mathcal{E}$ are eliminated

(*) $\mathcal{E}$ is always non-empty
(as it contains the infallible expert)

Hence, # mistakes $\leq \log_2 n$.

General case: Imperfect Experts

Compare the algorithm's performance.

$m_i^{(t)}$ – i – expert
t – time

$\in \{0,1\}$ 1, if expert i gets it wrong
0 otherwise

$$M^{(t)} = (m_1^{(t)}, m_2^{(t)}, \ldots m_n^{(t)})$$

① $\sum_{t=1}^{T} \min_{i \in [n]} m_i^{(t)}$ – infeasible to compare against $\Omega(T)$ additive diff

② $\min_{i \in [n]} \sum_{t=1}^{T} m_i^{(t)}$ $\rightarrow$ best expert w/ hindsight

---

Potential Approaches
- Act according to the majority
- Figure out who is a "good" expert in the first few rounds, and then act accg to that expert.

Proposed Alg: Same as Infallible case except lower wt of expert instead of making it zero.

Weighted Majority (WM_η) (Parameter η
$\eta \in (0, 1/2]$)

(1) Initialization:

For each expert $i \in [n]$

$$w_i^{(1)} \leftarrow 1$$

(2) For $t \leftarrow 1$ to T.

a. Make a decision based on weighted majority of experts' predictions using weights $w^{(t)} = (w_1^{(t)}, w_2^{(t)}, \ldots, w_n^{(t)})$

b. Observe the action.

c. Update the weights of all experts who predicted wrongly

$$w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot \frac{1}{2}$$

$$w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot (1-\eta)$$

$m_i^{(t)} = \mathbb{1}[\text{expert } i \text{ made a mistake at time } t]$

$$n_i^{(t)} = \sum_{s=1}^{t} m_i^{(s)}$$

$n^{(t)} =$ # mistakes made by WM till time t.

Thm: After T steps, for every expert i:

$$n^{(T)} \le 2.4\left(n_i^{(T)} + \log n\right)$$

(in particular, the above holds even w.r.t best expert $i^* = \arg\min_i n_i^{(T)}$)

Proof: Potential fn:

$$\Phi^{(t)} \triangleq \sum_{i=1}^{n} \omega_i^{(t)}$$

Initial: $\Phi^{(1)} = n$

Each time WM makes a mistake.

$$\Phi^{(t+1)} \le \Phi^{(t)}\left(\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\right)$$

$$\le \frac{3}{4}\cdot\Phi^{(t)}$$

$$\Phi^{(T+1)} \le \Phi^{(1)}\cdot\left(\frac{3}{4}\right)^{n^{(T)}} = n\cdot\left(\frac{3}{4}\right)^{n^{(T)}}$$

$$\Phi^{(T+1)} \ge \omega_i^{(T+1)} = \omega_i^{(1)}\cdot\left(\frac{1}{2}\right)^{n_i^{(T)}} = \left(\frac{1}{2}\right)^{n_i^{(T)}}$$

$$\left(\frac{1}{2}\right)^{n_i^{(T)}} \le n\cdot\left(\frac{3}{4}\right)^{n(T)}$$

$$n(T)\cdot\log\left(\frac{4}{3}\right) - \log n \le n_i^{(T)} \quad (\log \text{ w.r.t } 2)$$

$$n^{(T)} \le \frac{n_i^{(T)} + \log n}{\log(4/3)} \cong 2.4\left(n_i^{(T)} + \log n\right) \quad \boxtimes$$

Qn: (1) Is the multiplicative loss of 2.4 avoidable

(2) Is the additive loss of $O(\log n)$ avoidable.

Address (1)

$$WM \longrightarrow WM_\eta$$

$$w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot (1-\eta)$$

Analysis of WM_η:

$$\Phi^{(t)} \triangleq \sum_{i \in [n]} w_i^{(t)}.$$

Initially: $\Phi^{(1)} = 1$

Each time, WM_η commits a mistake

$$\Phi^{(t+1)} \leq \Phi^{(t)}\left(\frac{1}{2} + \frac{1}{2}(1-\eta)\right)$$

$$= \Phi^{(t)}\left(1 - \frac{\eta}{2}\right)$$

Hence, at time T

$$\Phi^{(T+1)} \leq \Phi^{(1)}\left(1-\frac{\eta}{2}\right)^{m^{(T)}}$$

On the other hand,

$$\Phi^{(t+1)} \geq w_i^{(t+1)} = w_i^{(t)} \cdot (1-\eta)^{n_i^{(T)}}$$

$$w_i^{(T+1)} \cdot (1-\eta)^{n_i^{(T)}} \leq n \cdot \left(1-\frac{\eta}{2}\right)^{n^{(T)}}$$

$$n^{(T)}\left(-\ln\left(1-\frac{\eta}{2}\right)\right) - \ln n \leq n_i^{(T)}\left(-\ln(1-\eta)\right)$$

$$n^{(T)} \leq n_i^{(T)}\left(\frac{\ln(1-\eta)}{\ln\left(1-\frac{\eta}{2}\right)}\right) + \frac{\ln n}{\left(-\ln\left(1-\frac{\eta}{2}\right)\right)}$$

$$\leq n_i^{(T)}\frac{(\eta+\eta^2)}{\eta/2} + \frac{\ln n}{\eta/2}$$

$$-\ln(1-x) \geq x$$
$$-\ln(1-x) \leq x+x^2 \quad \text{if } x \in (0, 1/2]$$

$$= 2(1+\eta)n_i^{(T)} + \frac{2\ln n}{\eta}$$

Thm: After T steps, for every expert i:

$$n^{(T)} \leq 2(1+\eta)\, n_i^{(T)} + \frac{2\log n}{\eta}$$

(In particular, the above holds even w.r.t best expert $i^* = \arg\min_i n_i^{(T)}$)

Qn: Is multiplicative factor of $2(1+\eta)$ unavoidable?

Yes, 2 is unavoidable if algorithm is deterministic

(not just for WM_n, but any det. algorithm)

— Get around the 2-multiplicative factor using randomness

Multiplicative Weight Update Method (MWUM)

More general,

Previously $m_i^{(t)} \in \{0,1\}$ Boolean value (win/loss)

Allow for non-Boolean rewards/losses

$$m_i^{(t)} \in [-1, 1]$$

$$M^{(t)} = (m_1^{(t)}, m_2^{(t)}, \ldots, m_n^{(t)}) \in [-1,1]^n$$

$$p^{(t)} = (p_1^{(t)}, p_2^{(t)}, \ldots, p_n^{(t)})$$

where each $p_i^{(t)}$ - prob w/ which alg acts accg to expert i at time t.

$\underline{MWUM_\eta}$ (Parameter: $\eta \in (0, \tfrac{1}{2}]$)

① Initialize: $\forall i \in [n]$, $w_i^{(1)} \leftarrow 1$

② For $t \leftarrow 1$ to T

(a) Choose a decision i according to the prob dist $p^{(t)} = (p_1^{(t)}, \ldots, p_n^{(t)})$

where $p_i^{(t)} = \dfrac{w_i^{(t)}}{\Phi^{(t)}}$ & $\Phi^{(t)} = \sum_{i \in [n]} w_i^{(t)}$

(b) Observe the costs $M^{(t)} = (m_1^{(t)}, \ldots, m_n^{(t)}) \in [-1, 1]^n$

(c) Update the weights

$$w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot (1 - m_i^{(t)} \cdot \eta)$$

—

$\ell(t) =$ Expected loss of $MWUM_\eta$ at time t

$$= \sum_{i \in [n]} m_i^{(t)} \cdot p_i^{(t)}$$

$$= \langle M^{(t)}, p^{(t)} \rangle$$

$L(t) =$ Expected loss at the end of time t

$$= \sum_{s=1}^{t} \ell(s)$$

Theorem: Assuming all costs $m_i^{(t)} \in [-1, 1]$ & $\eta \in (0, 1/2]$, then for any expert i

$$L(T) = \sum_{t=1}^{T} \langle M^{(t)}, p^{(t)} \rangle \le \sum_{t=1}^{T} m_i^{(t)} + \eta \sum_{t=1}^{T} |m_i^{(t)}| + \frac{\ln n}{\eta}$$

Take away: Randomness gets rid of the factor of 2.