6 Nov 2024

The HEDGE Algorithm.

Recall: We had n predictors and T questions with answers $y_1, \dots, y_T$.

$$
\begin{array}{c|cccc}
\text{Question} & 1 & 2 & \dots & T \\
\hline
\text{Predictor } 1 & 1 & 0 & & \\
2 & 0 & 1 & & \\
\vdots & 0 & 0 & & \\
 & 1 & 1 & & \\
n & 0 & 0 & &
\end{array}
$$

penalty for mistake

no penalty for answering correctly.

Let l_t be column t of this matrix.

In the "best expert" problem we generalize this to $l_t \in [0,1]^n$

In round t:

i. Algorithm selects probability distribution $p_t \in \Delta(n) = \left\{ \vec{p} \;\middle|\; \sum_{i=1}^{n} p_i = 1, \; p_i \geq 0 \right\}$

ii. loss vector $l_t \in [0,1]^n$ revealed.

iii. Algorithm suffers loss $\langle l_t, p_t \rangle$.

The HEDGE$_\varepsilon$ Algorithm

In round $t \in [T]$:

1. Let $L_{t-1} = \ell_1 + \ell_2 + \dots + \ell_{t-1}$ $(= 0 \text{ if } t = 1)$

2. For all $i \in [n]$ let

$$w_t(i) = (1-\varepsilon)^{L_{t-1,i}}$$

$$W_t = \sum_{i=1}^{n} w_t(i)$$

$$p_t = \frac{1}{W_t} \vec{w}_t$$

$0 \leq \ell \leq 1$
$\Downarrow$
$(1-\varepsilon)^{\ell} \leq 1 - \varepsilon \ell$

$$W_{t+1} = \sum_{i=1}^{n} w_{t+1}(i) = \sum_{i=1}^{n} w_t(i) \cdot (1-\varepsilon)^{\ell_{t,i}}$$

$$\leq \sum_{i=1}^{n} w_t(i) \cdot (1 - \varepsilon \ell_{t,i})$$

$$= W_t \sum_{i=1}^{n} p_t(i)(1 - \varepsilon \ell_{t,i})$$

$$= W_t \sum_{i=1}^{n} p_t(i) - \varepsilon W_t \sum_{i=1}^{n} p_t(i) \ell_t(i)$$

$$= W_t (1 - \varepsilon \langle \ell_t, p_t \rangle)$$

$$\ln W_{t+1} \leq \ln W_t + \ln(1 - \varepsilon \langle \ell_t, p_t \rangle)$$

$$\leq \ln W_t - \varepsilon \langle \ell_t, p_t \rangle$$

$$\ln W_{T+1} \le \ln W_0 - \varepsilon \sum_{t=1}^{T} \langle \ell_t, p_t \rangle$$

$$= \ln(n) - \varepsilon \sum_{t=1}^{T} \langle \ell_t, p_t \rangle$$

If i is any expert

$$W_{T+1} > w_{t+1}(i) = (1-\varepsilon)^{L_{t,i}}$$

$$\ln W_{T+1} > \ln(1-\varepsilon) \cdot L_{t,i}$$

$$\ln(1-\varepsilon) L_{t,i} < \ln(n) - \varepsilon \sum_{t=1}^{T} \langle \ell_t, p_t \rangle$$

$$\varepsilon \sum_{t=1}^{T} \langle \ell_t, p_t \rangle < \ln(n) - \ln(1-\varepsilon) \cdot L_{t,i}$$

$$\boxed{\sum_{t=1}^{T} \langle \ell_t, p_t \rangle} < \frac{\ln(n)}{\varepsilon} + \frac{1}{\varepsilon} \ln\left(\frac{1}{1-\varepsilon}\right) \cdot L_{t,i}$$

$$< (1+\varepsilon) \boxed{L_{t,i}} + \frac{\ln n}{\varepsilon}$$

Algorithm's loss

expert i's loss

$$\text{Regret}(T) = \sum_{t=1}^{T} \langle \ell_t, p_t \rangle - \min_i \{ L_{t,i} \}.$$

Hedge_ε has

$$\text{Regret}(T) < \varepsilon \cdot \min_i \{ L_{t,i} \} + \frac{\ln n}{\varepsilon}$$

$$< \varepsilon T + \frac{\ln n}{\varepsilon}$$

$$= 2\sqrt{T \ln n} \quad \left(\text{set } \varepsilon = \sqrt{\frac{\ln n}{T}} \right)$$

The Doubling Trick

Initialize $L^*_{est} = 4 \ln(n)$.

repeat:

  while $\min_i \{ L_{t,i} \} \leq L^*_{est}$:

    play Hedge_ε where $\varepsilon = \sqrt{\frac{\ln n}{L^*_{est}}}$.

  endwhile

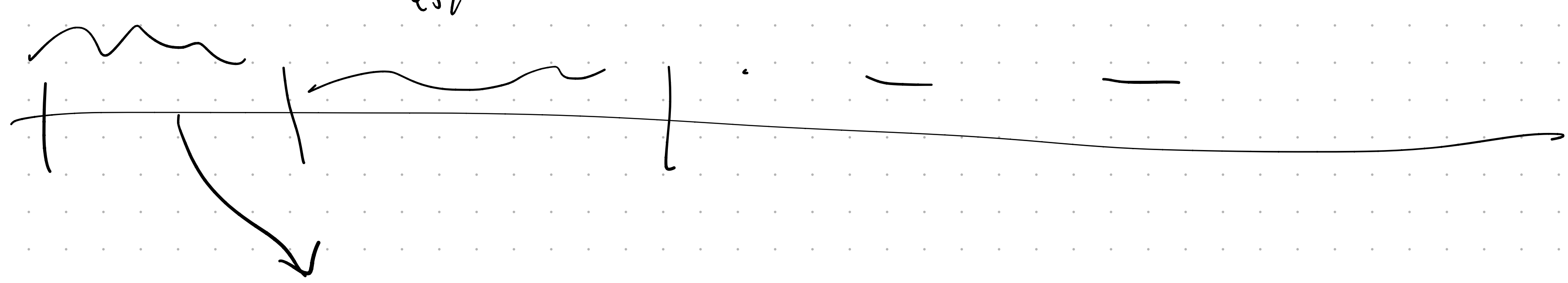
  Double the value of L^*_{est}

and repeat the while-loop with this new value.

(Reinitialize Hedge$_\varepsilon$ with $\emptyset$ initial loss for all experts every time.)

Timeline:

$L^*_{est} = 4\ln(n)$ $\quad$ $L^*_{est} = 8\ln(n)$



Regret: $\leq 2\sqrt{L^*_{est}\ln(n)}$

$$\text{Regret}(T) \leq O\left(\sqrt{L^{final}_{est}\ \ln(n)}\right).$$

$$\leq O\left(\sqrt{\min\{L_{T,i}\}\cdot \ln n}\right)$$