Expert Advice Scenario w/ N experts

For $t = 1, 2, \ldots$:

- learner observes the predictions of all N experts $b_{t,i} \in \{0,1\}$
- learner makes a prediction $a_t \in \{0,1\}$
- learner observes actual outcome $y_t \in \{0,1\}$

Under realizability assumption (i.e., $\exists$ some expert who never makes mistakes)

Follow the Leader: $\le N-1$ mistakes

Halving (Majority): $\le \lfloor \log_2 N \rfloor$ mistakes

"mistake bounds"

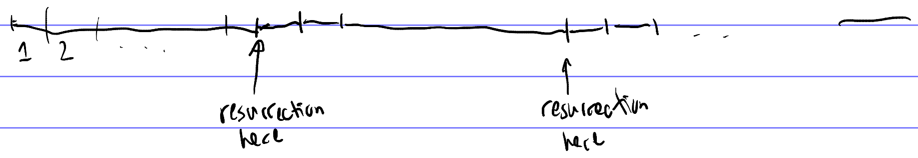
Lower bound: (In the EAS w/ N experts, realizability)
For any learner, there is a sequence of experts' predictions and true outcomes s.t. the learner makes $\lfloor \log_2 N \rfloor$ mistakes.

Change the scenario / assumptions to better match reality

Define "K-mistake assumption": At least one expert makes at most K mistakes.

A. Halving w/ Resurrection

- Same as Halving (copy most popular prediction among alive experts), but when all experts are dead, immediately "declare" that they are all "alive".



In a "block" of days before a resurrection, even the "best" expert made ≥ 1 mistake.

So $\le K$ resurrections.

So within these blocks, same argument as in analysis of Halving shows that learner makes $\le \lfloor \log_2 N \rfloor + 1$ mistake
"to kill off last expert"

\# mistakes of learner:

$$\le \underbrace{\#\text{ resurrections}}_{\le K} \times \underbrace{\#\text{ mistakes in block preceding resurrection}}_{\le \lfloor \log_2 N \rfloor + 1} + \underbrace{\#\text{ mistakes in block after last resurrection}}_{\le \lfloor \log_2 N \rfloor}$$

Thm In EAS w/ N experts K-mistake assump., HwR makes $\le (\lfloor \log_2 N \rfloor + 1) \cdot K + \lfloor \log_2 N \rfloor$ mistakes.

Lower bound In EAS w/ $N=2$ experts K-mistake assump., for every learner, there's a seq of experts' predictions and outcomes s.t. learner makes $\ge 2K+1$ mistakes.

Pf $N=2$ experts
Expert 0: always says 0
Expert 1: " " 1

For each of $2K+1$ days, adversary chooses outcome to be opposite of learner's prediction.
~~At least~~ one of the experts makes $\le K$ mistakes.

(Can combine w/ previous lower bound to say
$\#\text{mistakes} \ge \max\{\lfloor \log_2 N \rfloor, 2K+1\}$)

Exercise: Improve to $\lfloor \log_2 N \rfloor + 2K$

Weighted Majority (WM) hyperparameter $\beta \in (0,1)$

Initialize: $W_{1,i} = 1$ for each expert $i \in [N] = \{1, 2, \dots, N\}$

For day $t = 1, 2, \dots$

- Get $b_{t,i} \in \{0,1\}$ $\forall i \in [N]$
- Compute, for $y \in \{0,1\}$

$$Z_{t,y} := \text{sum of } W_{t,i} \text{ for all experts } i \text{ s.t. } b_{t,i} = y$$

- Predict

$$a_t = \begin{cases} 1 & \text{if } Z_{t,1} > Z_{t,0} \\ 0 & \text{o.w.} \end{cases}$$

- Get y_t
- Update weights

$$W_{t+1,i} := \begin{cases} W_{t,i} & \text{if } b_{t,i} = y_t \\ \beta W_{t,i} & \text{if } b_{t,i} \neq y_t \end{cases}$$

Analysis: Z_t = total "weight" at start of day t

$$= \sum_{i=1}^{N} W_{t,i}$$

$$= Z_{t,0} + Z_{t,1}$$

Suppose WM makes mistake on day t.

$$\alpha_t = \underbrace{\frac{Z_{t,1-y_t}}{Z_t}}$$ fraction of total weight that is wrong

$$\geq \frac{1}{2}$$

$$Z_{t+1} = Z_{t,y_t} + \beta Z_{t,1-y_t}$$

$$= (1-\alpha_t) Z_t + \beta \cdot \alpha_t Z_t$$

$$= (1-(1-\beta)\alpha_t) Z_t$$

$$\leq \left(1-(1-\beta)\cdot\tfrac{1}{2}\right) Z_t \qquad \beta = \tfrac{1}{2}$$

$$= \tfrac{3}{4} Z_t$$

So if M mistakes occur in t days,

$$Z_{t+1} \leq \left(\tfrac{3}{4}\right)^M \cdot Z_1 = \left(\tfrac{3}{4}\right)^M \cdot N$$

By K-mistake assumption, some expert has $W_{t+1,i} \geq 2^{-K}$

$$2^{-K} \leq Z_{t+1} \leq \left(\tfrac{3}{4}\right)^M \cdot N$$

$$-K \leq M \log_2 \tfrac{3}{4} + \log_2 N$$

$$M \leq \frac{1}{\log_2 4/3} \cdot (K + \log N) \approx 2.41(K + \log_2 N)$$

β	Mistake bound
$\frac{1}{2}$	$2.41K + 3.48 \ln N$
$\frac{2}{3}$	$2.22K + 5.48 \ln N$
$\frac{3}{4}$	$2.15K + 7.49 \ln N$
$\vdots$	