$\boxed{\text{6 Sep 2024}}$ Randomized Multiplicative Weights (Hedge)

<u>Recall:</u> WEIGHTED MAJ (ε) is deterministic $0 < \varepsilon < \frac{1}{2}$

and its mistake bound is

$$m_{ALG} < \frac{2\ln(K)}{\varepsilon} + 2(1+\varepsilon)\, m_{OPT}$$

<u>TODAY</u> A randomized alg (Hedge) that saves a factor of 2.

- K experts, $1, \dots, K$.
- At each time t:
  - ALG selects probab. distrib. over $[K]$, p_t.
  - ADV selects loss vector, $l_t \in [0,1]^K$.

p_t, l_t may depend on $\{(p_s, l_s) \mid s = 1, \dots, t-1\}$

- Loss of expert i is $l_t(i)$.

Algorithm's loss is $\langle l_t, p_t \rangle = \sum_{i=1}^{K} l_t(i)\, p_t(i)$.

Loss matrix

$$\begin{array}{cl} \begin{array}{ccccc} l_1 & l_2 & l_3 & \cdots & l_T \end{array} & \\ \left[\begin{array}{ccccc} 0 & 0 & 1 & & \\ 1 & 1 & 0 & \cdots & \\ 1 & 0 & 0 & & \\ 0 & 1 & 0 & & \end{array}\right] & \begin{array}{l} \leftarrow \text{expert 1} \\ \leftarrow \text{expert 2} \\ \vdots \\ \leftarrow \text{expert 4} \end{array} \end{array}$$

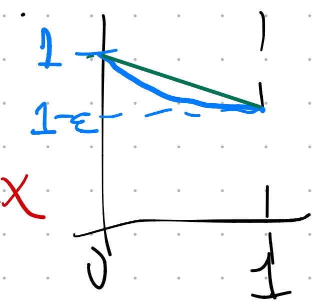
Define $L_t = \sum_{s=1}^{t} l_s$.

HEDGE(ε): for $t = 1, 2, \dots, T$:

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

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

$$p_t(i) = w_t(i) / W_t$$



$0 \le x \le 1$

$(1-\varepsilon)^x \le 1 - \varepsilon x$

Analysis of HEDGE.

$$w_{t+1}(i) = (1-\varepsilon)^{L_t(i)} = (1-\varepsilon)^{L_{t-1}(i) + \ell_t(i)}$$
$$= (1-\varepsilon)^{\ell_t(i)} \cdot w_t(i)$$
$$\le \left[1 - \varepsilon \ell_t(i)\right] \cdot w_t(i).$$

$$W_{t+1} = \sum_{i=1}^{K} w_{t+1}(i) \le \sum_{i=1}^{K} \left[1 - \varepsilon \ell_t(i)\right] \underbrace{w_t(i)}_{= p_t(i) W_t}$$
$$= \underbrace{\sum_{i=1}^{K} w_t(i)}_{= W_t} - \varepsilon \sum_{i=1}^{K} \ell_t(i) p_t(i) W_t$$
$$= W_t \cdot \left\{1 - \varepsilon \langle \ell_t, p_t \rangle\right\}$$

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

Inductively,

$$\ln W_{T+1} \le \underbrace{\ln W_0}_{= \ln K} - \varepsilon \underbrace{\sum_{t=1}^{T} \langle \ell_t, p_t \rangle}_{= \text{ALG cumulative loss.}}$$

If i^* is the best expert, meaning

$$L_T(i^*) = \min_{i \in [K]} \{L_T(i)\}$$

then

$$W_{T+1} = \sum_{i=1}^{K} (1-\varepsilon)^{L_T(i)}$$

$$> (1-\varepsilon)^{L_T(i^*)}$$

$$\ln W_{T+1} > \ln(1-\varepsilon) \cdot L_T(i^*)$$

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

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

$$\sum_{t=1}^{T} \langle \ell_t, p_t \rangle < \frac{\ln K}{\varepsilon} + \frac{-\ln(1-\varepsilon)}{\varepsilon} \cdot L_T(i^*)$$

$$< \frac{\ln K}{\varepsilon} + \frac{\varepsilon + \varepsilon^2}{\varepsilon} \cdot L_T(i^*)$$

$$= \ln(K)/\varepsilon + (1+\varepsilon) L_T(i^*)$$

<u>Definition</u>, (Regret)

$$R^{ALG}(T) = \underbrace{\sum_{t=1}^{T} \langle \ell_t, p_t^{ALG} \rangle}_{\text{ALG loss}} - \underbrace{L_T(i^*)}_{\text{Best expert's loss}}$$

For Hedge,

$$R^{HEDGE}(T) < \frac{\ln K}{\varepsilon} + \varepsilon L_T(i^*)$$

$$\leq \frac{\ln K}{\varepsilon} + \varepsilon T$$

Choose $\varepsilon = \sqrt{\frac{2 \ln K}{T}}$ ...

$$R^{HEDGE}(T) < \sqrt{2(\ln K)T} .$$

Or choose ε using "doubling trick":

At time t I don't yet know $L_T(i^*)$ but I do know

$$\min_{i \in [K]} \{ L_{t-1}(i) \}$$ ← loss of current best expert

Start by assuming $L_T(i^*)$ will be ≤ 1.

Run epochs $0, 1, 2, \ldots$.

In epoch j we assume $L_T(i^*)$ will be $\leq 2^j$ and set ε to minimize

$$\frac{\ln K}{\varepsilon} + \varepsilon \cdot 2^j \Rightarrow \varepsilon_j = \sqrt{\frac{2 \ln K}{2^j}} .$$

Epoch j ends when $\min_i \{L_{t-1}(i)\} > 2^j$.

Then re-initialize Hedge and begin epoch $j+1$.

$$\text{Regret}(\text{epoch } j) \leq \frac{\ln K}{\varepsilon_j} + \varepsilon_j 2^j$$

$$= \sqrt{2(\ln K) 2^j}$$

If we end in epoch J,

$$\text{Regret} \leq \sum_{j=0}^{J} \sqrt{2(\ln K) 2^j}$$

$$= \sqrt{2(\ln K)} \sum_{j=0}^{J} 2^{j/2} = O\left(\sqrt{2(\ln K) 2^J}\right)$$

$$= O\left(\sqrt{2(\ln K) L_T(i^*)}\right)$$