Learning (and Sleeping) with Experts & Bandits

Online learning: ML framework based on sequential data (stocks, language, weather, ... most things)
(has a certain "flavor" I want to communicate)

PROBLEM On day $t = 1, \ldots, T$

("ONLINE ALLOCATION")
- choose $p_t \in \Delta_{N-1}$ ← prob simplex on $[N]$ over "experts"
- sample $i_t \sim p_t$ $i_t \in [N]$

crucial step →
- nature reveals $\ell_t \in \mathbb{R}^N_+$
- learner incurs loss ℓ_{t,i_t} (in expectation, $\langle p_t, \ell_t \rangle$)

Interpretation / setup: You have N experts
Every day, they make a prediction, and each has "loss"/error $\ell_{t,i}$

Learner knows how experts have behaved in the past
maintains distribution p_t

GOAL minimize $\sum_{t=1}^{T} \ell_{t,i_t} - \min_{i \in [N]} \sum_{t=1}^{T} \ell_{t,i} = R_T$

in expectation, $\mathop{\mathbb{E}}_{i_t \sim p_t}(R_T) = \sum_{t=1}^{T} \langle \ell_t, p_t \rangle - \min_{i \in [N]} \sum_{t=1}^{T} \ell_{t,i}$

Algorithm: HEDGE (MWU)

Maintain $\eta > 0$. Initialize $w_1 = (1, \dots, 1)$ (N "weights")

On day $t = 1, \dots, T$:

- choose $p_t = \frac{w_t}{\|w_t\|_1} = \frac{w_t}{\sum w_{t,i}}$
- choose $i_t \sim p_t$
- upon recieving ℓ_t

$$\forall i \in [N] \quad w_{t+1,i} = w_{t,i} \exp(-\eta \ell_{t,i})$$

Thm Using HEDGE, with $\ell_t \in [0,1]^N$, $\mathbb{E}(R_T) \le \sqrt{2T \log N}$ (Corollary)

Thm $\sum_{t=1}^{T} \langle \ell_t, p_t \rangle - \min_{i \in [N]} \sum_{t=1}^{T} \ell_{t,i} \le \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \langle \ell_t^2, p_t \rangle$

(so average, per-round regret decays like $\frac{1}{\sqrt{T}}$) $H(p) = \sum p_i \ln 1/p_i$ entropy of p

Pf observe that $p_t = \min_{p \in \Delta^{N-1}} \left(\sum_{s=1}^{t-1} \langle \ell_s, p \rangle - \frac{H(p)}{\eta} \right)$

Standard proof: let $Z_t = \sum_{i=1}^{N} w_{t,i}$ "potential function"

Observe, $\forall i$, $Z_{T+1} > w_{T,i} = \exp\left(-\eta \sum_{t=1}^{T} \ell_{t,i}\right)$

$$\frac{Z_{t+1}}{Z_t} \le \exp\left(-\eta \langle \ell_t, p_t \rangle + \frac{\eta^2}{2} \langle \ell_t^2, p_t \rangle\right)$$

$$\Rightarrow \ln(Z_{T+1}) - \ln(Z_1) \le -\eta \sum_{t=1}^{T} \langle \ell_t, p_t \rangle + \frac{\eta^2}{2} \sum_{t=1}^{T} \langle \ell_t^2, p_t \rangle$$

link inequalities, rearrange □

PROBLEM on day $t = 1, \dots, T$

- choose $p_t \in \Delta_{N-1}$
- sample $i_t \sim p_t$
- nature reveals ONLY ℓ_{t,i_t}
- incur loss ℓ_{t,i_t}

"MULTI-ARMED BANDITS"

partial information

the only difference

Minimize $\sum \ell_{t,i_t} - \min_{i \in [N]} \sum \ell_{t,i}$

SAME REGRET

ALGORITHM (explore + exploit) w/ exponential weights, EXP3)

Maintain $\eta > 0$. $w_1 = (1, \dots, 1)$

On day $t = 1, \dots, T$:

- choose $p_t = w_t / \|w_t\|_1$
- choose $i_t \sim p_t$

$$w_{t,i} = \begin{cases} w_{t,i} & \text{for } i \neq i_t \\ w_{t,i} \exp\left(-\eta \ell_{t,i} / p_{t,i}\right) & \text{for } i = i_t \end{cases}$$

↑ inverse probability weighting

Idea: apply Hedge guarantee with loss vector

$$\tilde{\ell}_{t,i} = \mathbb{1}(i = i_t)\, \ell_{t,i} / p_{t,i}$$

$\Rightarrow$

$$\forall i \quad \sum_{t=1}^{T} \langle \tilde{\ell}_t, p_t \rangle \leq \sum_{t=1}^{T} \tilde{\ell}_{t,i} + \frac{\ln k}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \langle \tilde{\ell}_t^2, p_t \rangle$$

Take expectation: $\left(\tilde{\ell}_{t,i} \rightsquigarrow \ell_{t,i} \qquad \tilde{\ell}_{t,i}^2 \rightsquigarrow \tilde{\ell}_{t,i}^2 / p_{t,i}\right)$

$$R_T \leq \frac{\ln k}{\eta} + \frac{\eta k}{2} \sum_{t=1}^{T} \underbrace{\mathbb{E} \langle \tilde{\ell}_t^2, p_t \rangle}_{\sum_{i=1}^{N} \ell_{t,i}^2}$$

Note: We can decouple "action" and "expert"

for $t = 1, \dots, T$: choose $p_t \in \Delta_{k-1}$ (k actions, N experts)

$\rightsquigarrow$ receive expert actions $b_{t,i} \in [k]$ for $i \in [N]$

learner chooses action $a_t \sim p_t$

learner receives loss ℓ_{t,a_t} $\quad \ell_t \in \mathbb{R}_+^k$

(in previous setup, $K = N$ and no expert advice received)

Algorithm (EXP4):

for $t = 1, \dots, T$:

- choose $p_{t,a} = \dfrac{\sum_{i=1}^{N} w_{t,i} \mathbb{1}(b_{t,i} = a)}{\sum_{i=1}^{N} w_{t,i}} \in \Delta_{k-1}$
- choose $a_t \sim p_t$, receive ℓ_{t,a_t}
- create $\tilde{\ell}_{t,a} = \mathbb{1}(a_t = a) \, \ell_{t,a_t} / p_{t,a} \quad \forall a \in [k]$
- update $w_{t+1,i} = w_{t,i} \exp(-\eta \, \tilde{\ell}_{t,b_{t,i}})$ $\quad i \in [N]$

Sleeping experts

a more insidious kind of partial information...

On day $t = 1, \ldots, T$

— choose $p_t \in \Delta_{N-1}$

— nature reveals ONLY $\ell_{t,i}$ for $i \in E_t \subseteq [N]$ ("awake experts", adversarially chosen)

— incur loss $\sum_{i \in E_t} p_{t,i} \ell_{t,i}$

Our approach: "sleepy hedge"

if $i \notin E_t$, keep weight unchanged

$i \in E_t$, update $w_{t+1,i} = w_{t,i} \exp\left(-\eta \left(\ell_{t,i} - \frac{1}{|E_t|} \sum_{j \in E_t} \ell_{t,j}\right)\right)$

Guarantee:

$$R_T \leq \sqrt{2KNT \log N} + \sum_{t=1}^{N} \mu_t \left(p_{t,i^*} - \mathbb{1}(i^* \in E_t)\right)$$

$(\ell_{t,i} - \mu_t)^2 \leq K \quad \forall t, i$

can "assume this away"