The Experts Problem

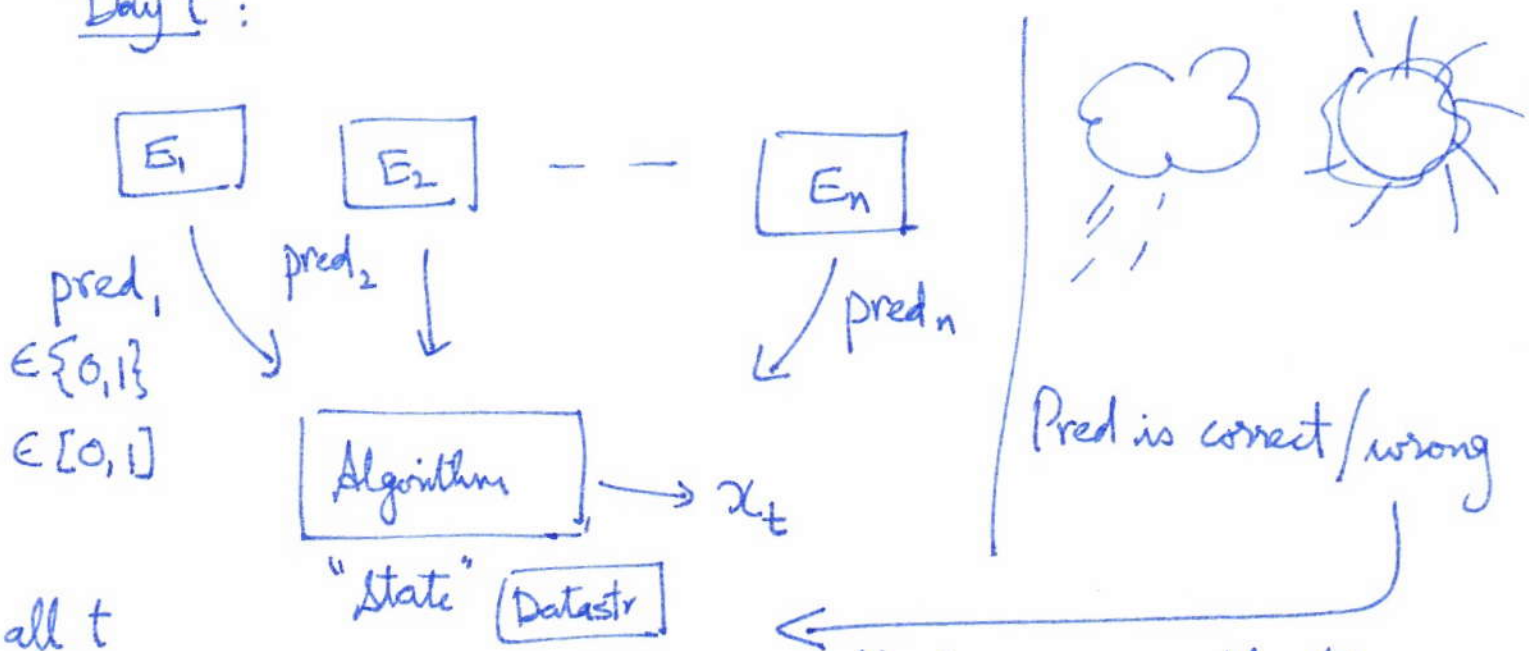
(Power of randomization in prediction)

The prediction problem: every day, we need to predict a Boolean.
(is it going to rain? is stock x going up?
will we see > x traffic on website today?)

There are n experts who make predictions? Who to follow?

Online setting

Day t:



For all t

1) Experts make predictions
2) Algorithm makes prediction
3) "Truth" is revealed
4) Algorithm incurs loss
5) Alg. does an update of data structure

Alg incurs penalty/loss

Alg updates data structure

$$Loss_t = \begin{cases} 0 & \text{if pred correct} \\ 1 & \text{if pred wrong} \end{cases}$$

$\rightarrow l_t(x_t) : [0,1] \rightarrow [0,1]$

Thm: T: time horizon
The loss of weighted majority is at most $2.5L + O(\log n)$

n = # experts
L = loss of the best expert

Proof: A potential argument. Define $Z(t) = \sum_{i \in [n]} w_t(i)$

Part 1: Argue that every time algorithm errs, potential drops significantly.

Part 2: If L is small, potential cannot be too small.

Conclusion: Hence, algorithm cannot err too many times.

Part 2: Consider best expert (say i^*)

$$w_T(i^*) = \frac{1}{2^L} \qquad Z(T) \geq \frac{1}{2^L}$$ (weights are non-negative)

Part 1: Suppose alg. made a mistake at time t.

At least half the "weight" got it wrong.

$$Z(t+1) = \sum_{\substack{i: \\ \text{wrong}}} w_{t+1}(i) + \sum_{\substack{i: \\ \text{right}}} w_{t+1}(i)$$

$$= \sum_{\substack{i: \\ \text{wrong}}} \frac{w_t(i)}{2} + \sum_{i: \text{right}} w_t(i) = \underbrace{\sum_{i \in [n]} w_t(i)}_{Z(t)} - \underbrace{\sum_{\substack{i: \\ \text{wrong}}} \frac{w_t(i)}{2}}_{\geq \frac{Z(t)}{4}}$$

(added/subtracted $\sum_{i: \text{wrong}} w_t(i)$)

$$\leq \left(\frac{3}{4}\right) Z(t)$$

How to state the guarantee?

If all the experts are bad, alg. has no hope.

If there exists a good expert, alg. should find that expert (alg. should incur small loss)

T : final round → time horizon

Loss function : $\ell_t : [n] \to \{0,1\}$ $\ell_t(i)$ is loss of i^{th} expert

The total loss of i^{th} expert $= \sum_{t=1}^{T} \ell_t(i)$

The loss of the best expert is $L := \min_{i \in [n]} \sum_{t=1}^{T} \ell_t(i)$

Benchmark: Optimal expert in hindsight.

Can algorithm achieve loss comparable to L?

Deterministic algorithm (Weighted Majority)

Data structure : array of weights, for each expert

$w_t : [n] \to (0,1]$ $\quad \forall i \in [n]$ $w_1(i) = 1$

At time t

1) Get prediction of all experts $\forall i$ $\text{pred}_t(i) \leftarrow \in \{0,1\}$
2) Compute total weight of experts predicting 0, total weight of experts predicting 1.
3) Predicts the majority weight (longer)

(Nature reveals)

4) For every expert i that is wrong $w_{t+1}(i) = w_t(i)/2$ $\quad w_t(i)(1-\eta)$

Weights are non-increasing. Hence $\forall t \quad Z(t+1) \leq Z(t)$

Let K denote loss of the algorithm

$$\left(\frac{1}{2}\right)^L \leq Z(T) \leq Z(1)\left(\frac{3}{4}\right)^K = n\left(\frac{3}{4}\right)^K$$

$$\Longrightarrow \quad L(\lg 1/2) \leq \lg n + K \lg\left(\frac{3}{4}\right)$$

$$K \leq \frac{L}{\lg(4/3)} + \frac{\lg n}{\lg(4/3)} \leq \underline{2.5}L + O(\lg n)$$

$K \leq (2+\eta)L + \left(\frac{\lg n}{\eta}\right)$ η is a parameter in $(0,1)$.

We get as close to $\underline{2}L + O(\lg n)$ as possible.

No deterministic algorithm can beat a loss of $2L$.

For any deterministic algorithm, experts and nature collude to force K to be large. Two experts $(n=2)$

At every round, one expert predicts 1 and the other predicts 0

Since algorithm is deterministic, nature can figure out ~~which expert algorithm~~ the algorithms prediction.

Nature behave the opposite.

Alg loss $= T$ Best expert loss $\leq \frac{T}{2}$

Randomized Weighted Majority

Maintain weights $w_t(i)$ (Parameter $\eta \in (0,1)$ learning rate)

At time t

1) Select expert $i \propto w_t(i)$
2) Follow prediction of chosen expert
3) After receiving loss:

Multiplicative Weights Update

$\forall i \quad w_{t+1}(i) = w_t(i)(1-\eta)^{\ell_t(i)} \left(w_t(i)\exp(-\eta \ell_t(i)) \right)$

In each time t, after alg. makes prediction, nature reveals "losses" for each expert $(\ldots, \ell_t(i), \ldots)$

Thm: $\mathbb{E}[\text{loss of alg}] \leq (1+\eta)L + O\left(\frac{\log n}{\eta}\right)$

Proof: $Z(t) = \sum_{i \in [n]} w_t(i)$

Part 2: $Z(T) \geq w_t(i^*) = \underbrace{w_t(1)}_{1}(1-\eta)^{\sum_{t=1}^{T} \ell_t(i^*)}$

$= (1-\eta)^L$

$Z(T) \geq \exp(-\eta L)$

Part 1: loss_t = loss of alg. at time t

$\mathbb{E}[\text{loss}_t] = \sum_{i \in [n]} \Pr[i \text{ is selected}]\, \ell_t(i) = \sum_{i \in [n]} \frac{w_t(i)}{Z(t)} \cdot \ell_t(i)$

$= \frac{1}{Z(t)} \sum_{i \in [n]} w_t(i)\ell_t(i)$

Potential reduction at time t

$$= Z(t) - Z(t+1) = \sum_{i\in[n]} w_t(i) - \sum_{i\in[n]} w_t(i)(1-\eta)^{l_t(i)}$$

$$(1-\eta)^{l_t(i)} \geq 1 - \eta l_t(i) \qquad (1-x)^n \geq 1-nx$$

$$Z(t) - Z(t+1) \leq \sum_{i\in[n]} w_t(i) - \sum_{i\in[n]} w_t(i)\left(1-\eta l_t(i)\right)$$

$$= \eta \sum_{i\in[n]} w_t(i) l_t(i) = \eta Z(t) \mathbb{E}[loss_t]$$

$$Z(t+1) \geq Z(t)\left(1-\eta \mathbb{E}[loss_t]\right)$$

$$Z(t) - Z(t+1) = \sum_{i\in[n]} w_t(i) - \sum_{i\in[n]} w_t(i) \exp\left(-\eta l_t(i)\right)$$

$$e^{-x} \leq 1 - x + \frac{x^2}{2}$$

$$\geq \sum_{i\in[n]} w_t(i) - \sum_{i\in[n]} w_t(i)\left(1 - \eta l_t(i) + \frac{(\eta l_t(i))^2}{2}\right) \qquad l_t(i) \leq 1$$

$$= \eta \sum_{i\in[n]} w_t(i) l_t(i) - \frac{\eta^2}{2} \sum_{i\in[n]} w_t(i) l_t(i)^2 \geq \eta \sum_{i\in[n]} w_t(i) l_t(i) - \frac{\eta^2}{2} \sum_{i\in[n]} w_t(i) l_t(i)$$

$$= \eta \mathbb{E}[loss_t] Z(t) - \frac{\eta^2}{2} \sum_{i\in[n]} w_t(i) l_t(i) = \eta \mathbb{E}[loss_t] Z(t) - \eta^2/2 \; \mathbb{E}[loss_t] \cdot Z(t)$$

$$Z(t) - Z(t+1) \geq Z(t)\left(\eta \mathbb{E}[\text{loss}_t] - \frac{\eta^2}{2}\mathbb{E}[\text{loss}_t]\right)$$

$$Z(t+1) \leq Z(t)\left(1 - \mathbb{E}[\text{loss}_t]\left(\eta - \frac{\eta^2}{2}\right)\right)$$

$$\leq Z(t)\exp\left(-\left(\eta - \frac{\eta^2}{2}\right)\mathbb{E}[\text{loss}_t]\right)$$

$$\exp(-\eta L) \leq Z(T) \leq \underbrace{Z(1)}_{n} \exp\left(-\mathbb{E}[\text{loss}]\left(\eta - \frac{\eta^2}{2}\right)\right)$$

$$-\eta L \leq \ln n - \left(\eta - \frac{\eta^2}{2}\right)\mathbb{E}[\text{loss}]$$

$$\mathbb{E}[\text{loss}] \leq \frac{\eta L}{(\eta - \eta^2/2)} + \frac{\ln n}{\eta - \eta^2/2}$$

$$\leq (1+\eta)L + O\left(\frac{\ln n}{\eta}\right) \qquad L + \eta L + \frac{\ln n}{\eta}$$

If $\eta = 1/\sqrt{T}$ $\quad T \geq L$ $\quad \eta T = \frac{\ln n}{\eta L}$ $\quad \eta = \sqrt{\frac{\ln n}{T}}$

$$\mathbb{E}[\text{loss}] \leq L + \frac{L}{\sqrt{T}} + \sqrt{\ln n} \times \sqrt{T} \qquad L \leq T$$

$$\leq L + O(\sqrt{T \ln n}) \quad \text{Sublinear}$$

$\mathbb{E}[\text{loss}]$ − best loss of expert $\leq O(\sqrt{T \ln n}) = o(T)$ → Regret