# "No-Regret" Learning

# Imagine that...

- On each day $t$, we must predict an outcome $y_t \in \{0,1\}$
- We have a vector $x_t \in \{0,1\}^n$ of features/advisors/experts we can use to help
- But now we assume nothing about $\langle x_t, y_t \rangle$
- No $x_t \sim P$, no $y_t = c(x_t)$
- An arbitrary, even adversarial sequence

# Detailed protocol

For $t = 1, \ldots T$:

- Nature chooses $\langle x_t, y_t \rangle$
- We are shown $x_t$ only
- We make prediction $\hat{y}_t$
- We are shown $y_t$

$\langle x_t, y_t \rangle$ may depend on all previous $\langle x, y \rangle$, our previous $\hat{y}$, current state our algo

Q: What could we possibly hope to do/say?

A: When there's no link between past and future, say something about the past.

- Let $x_t^i$ = $i^{th}$ bit of $x_t$
- $\ell_t^i = \mathbb{I}[x_t^i \neq y_t]$
- $L_T^i = \sum_{t=1}^{T} \ell_t^i$ = total loss of $i$

---

Idea: On any sequence, let's try to compete with $\min_i \{L^i\}$ in hindsight.

# Multiplicative Weights Algo

- Initialize $w_1^i = 1, p_1^i = 1/n \ \forall i$
- For $t = 1, \dots T$:
  - receive $x_t$
  - draw $j \sim p_t$, predict $\hat{y}_t = x_t^j$
  - receive $y_t$, compute the $\ell_t^i$
  - $\forall i$: $w_{t+1}^i \leftarrow w_t^i (1 - \eta \ell_t^i)$
  - $\forall i$: $p_{t+1}^i \leftarrow w_{t+1}^i / Z_{t+1}$, where

$$Z_{t+1} = \sum_i w_{t+1}^i$$

- Let $\ell_t^A = E_{p_t}[\mathbb{I}[\hat{y}_t \neq y_t]]$

$$L_T^A = \sum_t \ell_t^A$$

## Idea of analysis

- $L_T^A$ large $\Rightarrow Z_{T+1}$ small ①

- But $Z_{T+1} \geq$ largest weight at $T$ ② (best $x^i$)

# Analysis

For any $t$:

$$Z_{t+1} = \sum_i w_{t+1}^i$$

$$= \sum_i w_t^i (1 - \eta \ell_t^i)$$

$$= \sum_i w_t^i - \sum_i w_t^i \eta \ell_t^i$$

$$= Z_t - \eta Z_t \sum_i \frac{w_t^i}{Z_t} \cdot \ell_t^i$$

$$= Z_t (1 - \eta \sum_i p_t^i \ell_t^i) \quad \ell_t^A$$

$$= Z_t (1 - \eta \ell_t^A)$$

$$\therefore Z_{T+1} = Z_1 \prod_{t=1}^{T} (1 - \eta \ell_t^A)$$

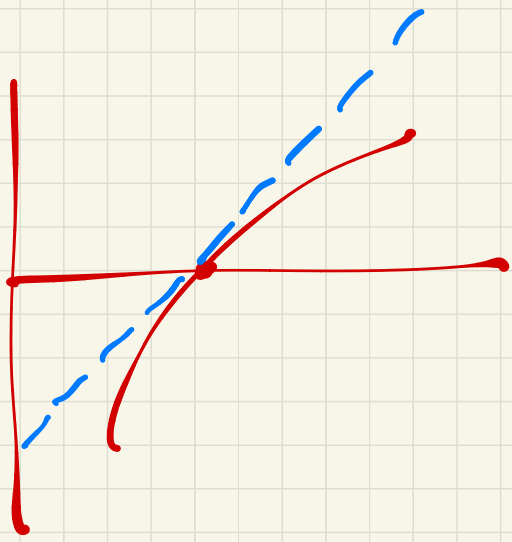
$$= n \cdot \prod_t (1 - \eta \ell_t^A)$$

$$\ln(Z_{T+1}) =$$

$$\ln(n) + \sum_t \ln(1-\eta l_t^A)$$

$$\leq \ln(n) - \eta \sum_t l_t^A \quad \left(\begin{array}{l}\ln(1-x) \\ \leq -x\end{array}\right)$$

$$= \ln(n) - \eta L_T^A \quad (1)$$



Now for any $i$:

$$Z_{T+1} \geq w_{T+1}^{i} \quad ②$$

$$\ln(Z_{T+1}) \geq \ln(w_{T+1}^{i})$$

$$= \ln\left(\prod_{t=1}^{T}(1-\eta \ell_t^i)\right)$$

$$= \sum_t \ln(1-\eta \ell_t^i)$$

$$\geq \sum_t \left(-\eta \ell_t^i - (\eta \ell_t^i)^2\right) \quad \left(\begin{array}{l} \ln(1-x) \geq \\ -x-x^2, \\ x \in [0, 1/2] \end{array}\right)$$

$$= -\eta \sum_t \ell_t^i - \eta^2 \sum_t (\ell_t^i)^2$$

$$= -\eta L_T^i - \eta^2 Q_T^i$$

So:

$$\ln(Z_{T+1})$$

$$-\eta L_T^i - \eta^2 Q_T^i \underset{\substack{\uparrow \\ \text{for } \underline{\text{any}}\ i}}{\leq} \cdot \leq \ln(n) - \eta L_T^A$$

$$L_T^A \leq \frac{\ln(n)}{\eta} + L_T^i + \eta Q_T^i$$

$$\leq \frac{\ln(n)}{\eta} + L_T^i + \eta T$$

$$\therefore L_T^A \leq \min_i \{L_T^i\} + \underbrace{\frac{\ln(n)}{\eta} + \eta T}_{\text{"regret"}}$$

Now set $\frac{\ln(n)}{\eta} = \eta T$

$$\eta^2 = \frac{\ln(n)}{T}, \eta = \sqrt{\frac{\ln(n)}{T}}$$

$\therefore$ both terms $= \eta T$

$$= \sqrt{\frac{\ln(n)}{T}} \cdot T = \sqrt{T \ln(n)}$$

Theorem For any sequence

$\langle x_t, y_t \rangle$,

$$L_T^A \le \min_i \{ L_T^i \} + 2\sqrt{T \ln(n)}.$$

$$\text{Or: } \frac{L_T^A}{T} \le \min_i \left\{ \frac{L_T^i}{T} \right\} + \underbrace{2\sqrt{\frac{\ln(n)}{T}}}_{\to 0}$$

"no regret"

# Discussion

- Sanity check: $n = 2^T$, all possible binary sequences
- If $\min_i \{L_T^i\}$ is bad, we are too
- Compare to building a model on top of $x^i$
- "Bookkeeping" vs. "learning"?

# Connection to PAC

- Suppose that each day there is $\langle z_t, y_t \rangle \sim P$ and $x_t^i = h(z_t)$
- Here $h \in \mathcal{H}$, $\mathcal{H}$ finite
- Recall uniform convergence:

$$m \sim \frac{\log|\mathcal{H}|}{\varepsilon^2} \text{ or } \varepsilon \sim \sqrt{\frac{\log|\mathcal{H}|}{m}}$$

- MW regret, $n = |\mathcal{H}|$:

$$\sim \sqrt{\frac{\log|\mathcal{H}|}{T}}$$ (but must enumerate)

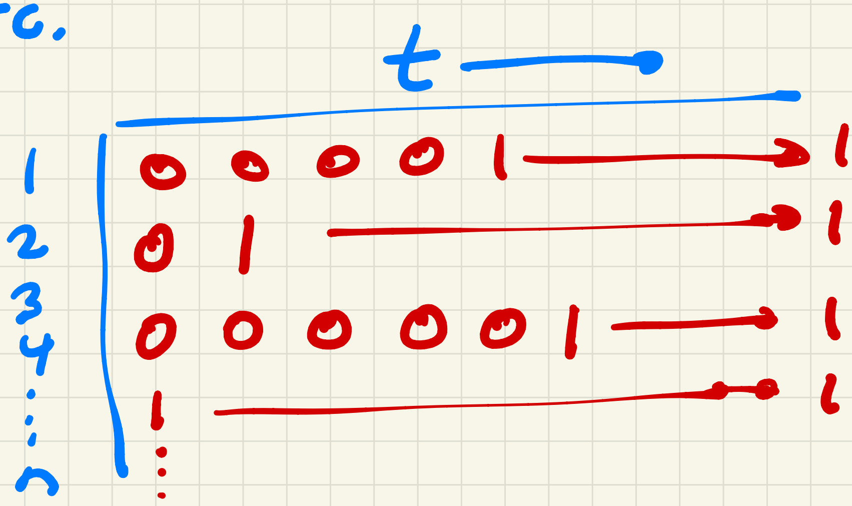
# More General Framework

- Before: $\ell_t^i = \mathbb{I}[x_t^i \neq y_t]$
- But MW analysis only looked at $\ell_t^i$
- So get rid of $x_t, y_t$
- Abstract actions $i \in \{1, \ldots n\}$
- Arbitrary $\ell_t^i$ = loss of $i$ on day/trial $t$

- MW just chooses an action $i$ each day $t$, then learns all $l_t^i$, pays $l_t^i$.
- Identical analysis/result
- $l_t^i$ can also be real-valued

# Lower Bounds on Regret

- Let $A$ be any algo
- At $t=1$, randomly choose
(A) $n/2$ actions to have loss $=0$,
(B) other $n/2$ have loss $=1$
- $E[\text{loss of } A @ t=1] = 1/2$
- At $t=2$, randomly choose
  $1/2$ of (A) to have loss $=0$,
  all others loss $=1$
- $E[\text{loss of } A @ t=2] \geq 1/2$
- etc.



- Can continue for $\log(n)$ steps with a 0-loss action
- But $E[\text{loss of } A] \geq \frac{1}{2}\log(n)$
- $\therefore$ total regret $\geq \frac{1}{2}\log(n)$

# Lower Bound #2

- $n=2$, losses will be $(0,1)$ or $(1,0)$
- Choose randomly for each step $t$
- For any distribution $(p, 1-p)$, $E[\text{loss}] =$

$$\frac{1}{2}(1-p) + \frac{1}{2}p = 1/2$$

- $E[\text{total loss of } A] = T/2$

- Also $L_T^1 + L_T^2 = T$

$$\begin{array}{cc} L_T^1 & L_T^2 \\ \downarrow & \downarrow \\ T/2 - y & T/2 + y \end{array}$$

- But $E[|y|] \sim \sqrt{T}$

$E[\text{regret of } A]$

# No-Regret Learning
# and
# Game Theory

# Game Theory Basics

- Two players with loss functions $\ell_1(i,j), \ell_2(i,j)$
- Player 1 chooses $i$, 2 chooses $j$
- One-shot, simultaneous move
- Players may randomize

| | R | P | S |
|---|---|---|---|
| R | 0 0 | 1 -1 | -1 1 |
| P | -1 1 | 0 0 | 1 -1 |
| S | 1 -1 | -1 1 | 0 0 |

# Safety Levels

- Imagine your opponent plays adversarially

Fact: Must exist values $V_1, V_2$ & distributions $P_1, P_2$ s.t.

- $\forall Q \quad \mathop{E}_{\substack{i \sim P_1 \\ j \sim Q}} \left[ \ell_1(i,j) \right] \le V_1$

- $\forall Q \quad \mathop{E}_{\substack{i \sim Q \\ j \sim P_2}} \left[ \ell_2(i,j) \right] \le V_2$

Now imagine the game is played repeatedly for T rounds, with 1 playing no-regret and 2 arbitrarily.

What happens?

<u>Theorem</u> Suppose player 1 uses an algo $A$ with cumulative regret $R(T)$ in $T$ steps, and player 2 plays arbitrarily. Then the total loss $L_T^1$ obeys:

$$\frac{L_T^1}{T} \leq V_1 + \frac{R(T)}{T}.$$

Similarly

$$\frac{L_T^2}{T} \leq V_2 + \frac{R(T)}{T}.$$

**Proof** Let $a_t^2$ be action of 2 at step $t$, and define:

$$\hat{P}_2(j) = \frac{1}{T}\sum_t \mathbb{I}[a_t^2 = j]$$

(empirical dist. of 2)

Then by 1's safety level,

$$\exists P_1 \text{ s.t. } \mathop{E}_{\substack{i \sim P_1 \\ j \sim \hat{P}_2}} [\ell_1(i,j)] \leq V_1$$

$\therefore \exists$ fixed $i^*$ s.t.

$$\underbrace{\mathop{E}_{j \sim \hat{P}_2} [\ell_1(i^*,j)] \leq V_1}$$

per-step loss of $i^*$ in hindsight

$\therefore$ total loss of $i^*$ in hindsight $\leq v_1 \cdot T$

$$\Rightarrow \frac{L_T'}{T} \leq v_1 + \frac{R(T)}{T}$$

# Zero-Sum Games

- $\forall i,j \quad \ell_1(i,j) = -\ell_2(i,j)$
- Define:

$$V_1^{min} \triangleq \max_{p_2} \min_{i} \mathbb{E}_{j \sim p_2}[\ell_1(i,j)]$$

(1 moves last)

$$V_1^{max} \triangleq \min_{p_1} \max_{j} \mathbb{E}_{i \sim p_1}[\ell_1(i,j)]$$

(1 moves first)

Have $V_1^{min} \leq V_1^{max}$, $V_2^{min} \leq V_2^{max}$

$$V_1^{min} = -V_2^{max}$$

$$V_2^{min} = -V_1^{max}$$

# Minimax Theorem:

$$V_1^{min} = V_1^{max}, \quad V_2^{min} = V_2^{max}.$$

Pf. Suppose for $\Rightarrow\Leftarrow$:

$$V_1^{max} = V_1^{min} + \gamma, \quad \gamma \gneq 0.$$

Let both players use a no-regret algo with $R = R(T)$.

Let $\hat{P}_1, \hat{P}_2$ be empirical distributions of algos.

Note $L_T^1 = -L_T^2$ always.

- $\hat{P}_2$ a possible choice for $P_2$ in $V_1^{min}$ $\Rightarrow$

total loss $\leq T V_1^{min}$ possible for 1

(total $\leq T V_2^{min}$ possible for 2)

$$L_T^1 \leq T V_1^{min} + R \quad (1)$$

$$L_T^2 \leq T V_2^{min} + R$$

$$L_T^2 \le T v_2^{min} + R$$

$$-L_T^1 \le T v_2^{min} + R$$

$$L_T^1 \ge -T v_2^{min} - R$$

$$L_T^1 \ge T v_1^{max} - R \quad (2)$$

$$T v_1^{max} - R \le L_T^1 \le T v_1^{min} + R$$

$$v_1^{max} - R/T \le v_1^{min} + R/T$$

$$v_1^{min} + \gamma - R/T \le v_1^{min} + R/T$$

For $R/T < \gamma/2$, $\Rightarrow\Leftarrow$.

(e.g. MW algo)

Furthermore:

Against $\hat{P}_2$, no $P_1$ can beat best $i$, but $\hat{P}_1$ almost as good

Against $\hat{P}_1$, no $P_2$ can beat best $j$, but $\hat{P}_2$ almost as good

$\therefore$ $(\hat{P}_1, \hat{P}_2)$ is within $R/T$ of minimax solution

$\equiv$ Nash equilibrium

<u>Important</u> Approximation quality depends only logarithmically on # of actions!