Last time

- PAC learning
- Gradient descent & SGD

Today

- online learning
  Multi-weight update alg
- applications including
  zero-sum games

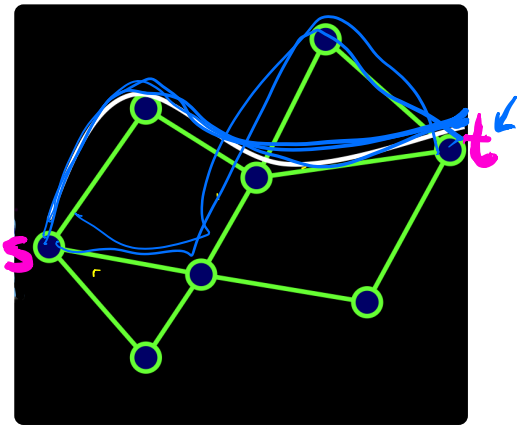
Figure credits

- Avrim Blum slides
- Emily Fox 446 slides
- book on Boosting by Schapire & Freund
- Nate Jensen, Gabrielle Cohn

$A = \{P_1, P_2, \ldots, P_m\}$

$R_1^t \; R_2^t \; \cdots \; R_m^t$

Online learning.



A : action set $|A| = n$

reward vector on day t $\quad r^t : A \to [-1, 1]$

$r^t(a)$: payoff on day t if you choose action a

Alg: on day t, chooses $a_t \in A$

knows only $r^1, r^2, \ldots, r^{t-1}$

$t = 1, \ldots, T$

alg $\xrightarrow{a_t}$ adversary $\xrightarrow{r^t}$ alg

$p^t(a)$: Prob choose action a on day t

adversary knows the prob distn over actions that you are choosing from

What can you hope to achieve?

Best possible outcome $\quad \sum_{t=1}^{T} \max_a r^t(a)$

$\sum_{a \in A} p^t(a) = 1$

This benchmark is way too strong.

$A = \{1, 2\}$

$p^t(1) + p^t(2) \le 0$

if $p^t(1) \ge \frac{1}{2}$ $\quad r^t(1) = -1, \; r^t(2) = 1$

≤ 0

o.w. $p^t(2) > \frac{1}{2}$ $\quad r^t(1) = 1, \; r^t(2) = -1$

benchmark T

combining expert advice or doing as well as best expert.

$$\text{Regret}(a^1, a^2, \ldots, a^T) = \max_{a \in A} \sum_{t=1}^{T} r^t(a) - \sum_{t=1}^{T} r^t(a_t)$$

best reward if you use same action every day

alg total reward.

on day t, need to choose $a \in A$.

"Follow the leader"

Claim: without randomization, impossible to do well

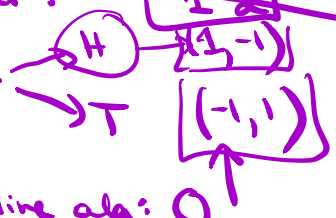
on day t $\quad r^t(a_t) = 0$, $\quad r^t(a) = 1$ $\quad a \neq a_t$ $\Longrightarrow$ alg total reward $= 0$.

n actions $\quad \sum_{t=1}^{T} \sum_{a \in A} r^t(a) = (n-1)T$

$\exists$ action w/ reward at least $\frac{(n-1)T}{n} = \left(1 - \frac{1}{n}\right)T \geq \frac{T}{2}$

What is the best we can hope for?

$n = 2$ toss a fair coin:



Expected reward of any online alg: 0

if pick action 1, $\#H\text{'s} - \#T\text{'s}$

2, $\#T\text{'s} - \#H\text{'s}$.

$$E\left(|\#H\text{'s} - \#T\text{'s}|\right) = \Theta(\sqrt{T})$$ ⟸ central limit theorem

Regret is $\Omega(\sqrt{T})$

$$\frac{\text{regret}}{T} = O\left(\frac{1}{\sqrt{T}}\right) \to 0$$

Thm: ∃ online alg with exp regret $2\sqrt{T \ln n}$

$$\max_a \sum_t r^t(a) - E(\text{reward of online alg}) \le 2\sqrt{T \ln n}$$

$$\frac{1}{T} E(\text{reward of online alg}) \ge \frac{1}{T} \max_a \sum_t r^t(a) - \frac{2\sqrt{T \ln n}}{T}$$

(max: total reward of best action)

$$2\sqrt{\frac{\ln n}{T}}$$

$-1 \le r^t(a) \le 1 \quad \forall a$

MWU algorithm

initialize $w^1(a) = 1 \quad \forall a$ — initialize all weights to 1

for $t = 1$ to T

pick action a with probability proportional to $w^t(a)$

given r^t, update weights.

$$w^{t+1}(a) = w^t(a) \cdot (1 + \eta \cdot r^t(a)) \quad (*)$$

learning rate parameter $\eta \in (0, \frac{1}{2})$

$$p^t(a) = \frac{w^t(a)}{\sum_{\tilde{a} \in A} w^t(\tilde{a})} \leftarrow \Gamma^t$$ sum of all wts at time t

$$E[\text{reward of MWU alg at time } t] = \sum_a p^t(a) r^t(a) = \frac{1}{\Gamma^t} \sum_a w^t(a) r^t(a) = \gamma_t$$

$$\Gamma^{t+1} = \sum_a w^t(a)(1 + \eta r^t(a)) = \Gamma^t + \eta \sum_a w^t(a) r^t(a)$$

$$\Gamma^t \gamma_t$$

Suppose a^* is best action in hindsight

$$\Gamma^{T+1} \ge w^{T+1}(a^*) = w^1(a^*) \prod_{t=1}^{T} (1 + \eta r^t(a^*)) = e^{\ln(1 + \eta r^t(a^*))}$$

$$\Gamma^{t+1} = \Gamma^t (1 + \eta \gamma_t) \le \Gamma^t e^{\eta \gamma_t} \quad \forall t \qquad 1 + x \le e^x$$

$$\Gamma^{T+1} \le \Gamma^T e^{\eta \gamma_T} \le \Gamma^1 e^{\sum \eta \gamma_t} = n e^{\sum_{t=1}^{T} \eta \gamma_t}$$

(b) — # actions

$\ln(1+x) \ge x - x^2$

$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \cdots$

$-\frac{x^2}{2}$

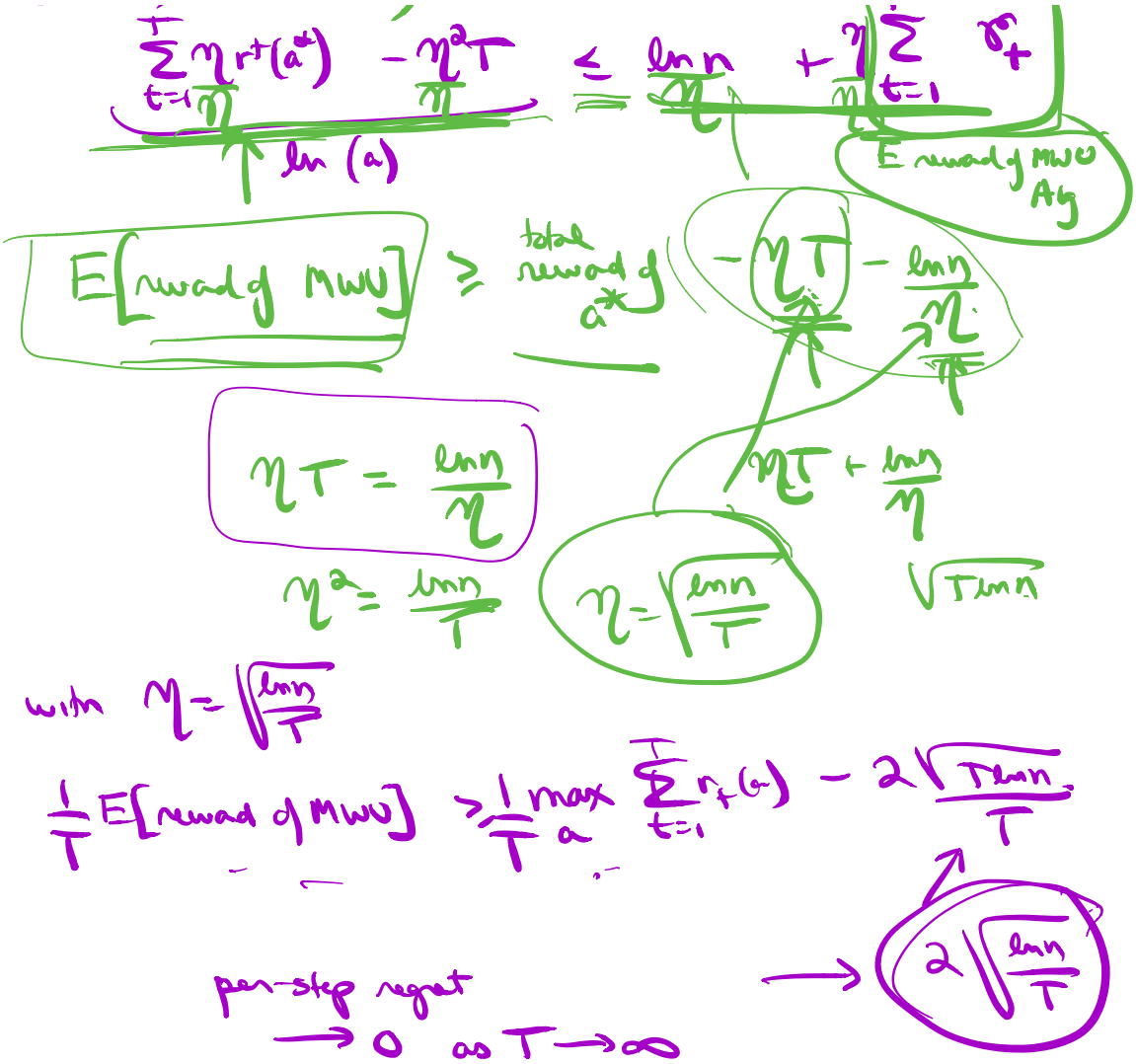
$$\ge e^{\sum_t \ln(1 + \eta r^t(a^*))}$$

$$\ge e^{\sum_{t=1}^{T} [\eta r^t(a^*) - \eta^2 r^t(a^*)^2]}$$

$$\ge e^{E[\sum \eta r^t(a^*)] - \eta^2 T}$$

(a)

$(a) \le (b) \quad \ln(a) \le \ln(b)$

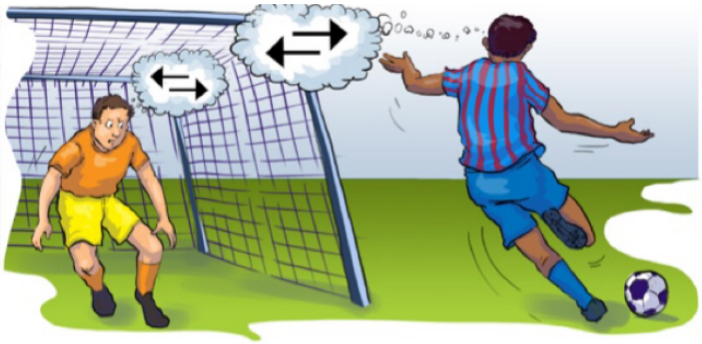


with $\eta = \sqrt{\frac{\ln n}{T}}$

$$\frac{1}{T} E[\text{reward of MWU}] \ge \frac{1}{T} \max_a \sum_{t=1}^{T} r_t(a) - 2\sqrt{\frac{T \ln n}{T}}$$

per-step regret $\to 0$ as $T \to \infty$ $\longrightarrow$ $2\sqrt{\frac{\ln n}{T}}$

2-player Zero-sum games



Penalty Kicks.

zero-sum means that one player's gain is other players loss.

col player. goalee

row player. kicker →	L	R
L	0	1
R	1	0

payoff to row player

rows (cols) are called "pure" strategies.

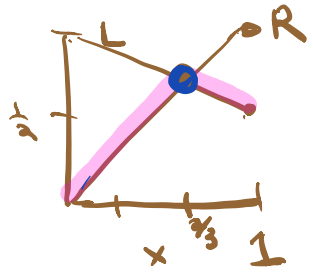
mixed strategy is prob distn over pure strategies.

goalee.

kicker	L	R
L (x)	1/2	1
R (1-x)	1	0

$V = 2/3$

What is opt mixed strategy for kicker if he has to go first.



$1 - .5\frac{2}{3} = \frac{2}{3}$

if goalee goes L,
exp loss = $.5x + 1 - x$
$= 1 - .5x$

if goalee goes R
exp loss = x

$1 - .5x = x$

$x = \frac{2}{3}$

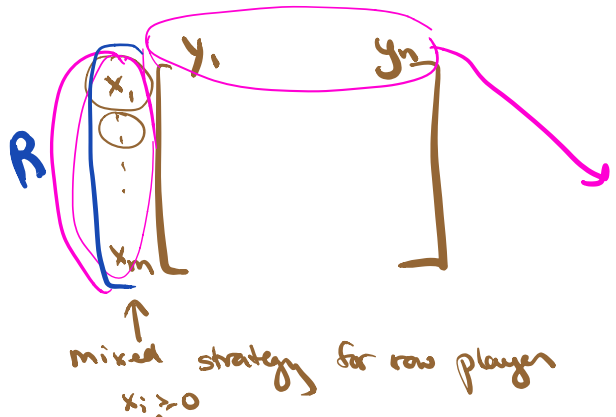
if kicker goes first, maximize loss of goalee

$$\max_x \text{ gain of kicker when goalee best responds to } x$$

minimax opt strategy for row players

zero sum game defined $m \times n$ matrix A

where a_{ij} is payoff to row player when he plays pure strategy i & col player plays pure strategy j



mixed strategy for col player

y_j: Pr(col player plays col j)

mixed strategy for row players

$x_i \geq 0$

$\sum_{i=1}^{m} x_i = 1$

$$E_{x,y}[\text{payoff to row player}] = \sum_{i=1}^{m}\sum_{j=1}^{n} \Pr(\text{Row-}i\ \text{Col-}j)\, a_{ij}$$

$\Pr(\text{Row-}i\ \text{Col-}j) = x_i \cdot y_j$

$$= \sum_i \sum_j x_i y_j a_{ij}$$

$$= x^T A y$$

if row player & chooses goes first $\vec{x}$

$\Rightarrow$ col player will choose

$$\max_x \left(\min_y x^T A y \right)$$

best response to x

$=$

if col player first & chooses y

$$\min_y \left(\max_x x^T A y \right)$$

x row player mixed str

y col play mixed str.

$=$ value game.

Portfolio Selection.

Stock	p_i^1	$r_i^1 = \frac{\text{closing price}}{\text{opening price}}$		r_i^2	
1	0.5	1.5		0.9	0.3
2	0	1		.1	0.1
3	0.5	1		0	0.4
4	0	0.5		0	0.2

\$1 $\quad 0.5 \cdot 1.5 + 0.5 \cdot 1 = \$1.25$

\$$W_0$ $\quad 1.25\, W_0$

p_i^t: fraction of your wealth that you put into stock i on day t.

r_i^t: $\frac{\text{closing price of stock } i \text{ on day } t}{\text{opening price}}$

$\frac{W_T}{W_0} = r_i^1 r_i^2 r_i^3 \cdots r_i^T$ $\quad W_0 r_i^1 r_i^2$

$\log\left(\frac{W_T}{W_0}\right) = \sum_{t=1}^{T} \log r_i^t$

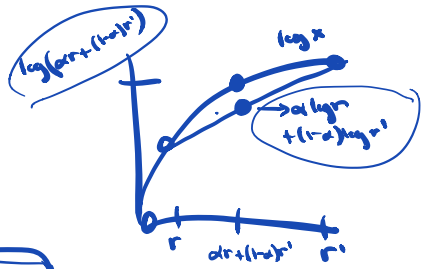
Let's pretend reward of stock i on day t is $\log r_i^t$

$$\sum_{t=1}^{T} \sum_i p_i^t \log r_i^t \geq \sum_{t=1}^{T} \log r_j^t - \varepsilon \quad \forall j$$

What we really care about:

$$\frac{W_T}{W_0} = \prod_{t=1}^{T} \left(\sum_i p_i^t r_i^t \right)$$

$$\log\left(\frac{W_T(\text{new})}{W_0}\right) = \sum_{t=1}^{T} \log\left(\sum_{i=1}^{n} p_i^t r_i^t\right)$$



$$\sum_{t=1}^{T} \log\left(\sum_{i=1}^{n} p_i^t r_i^t\right) \geq \sum_{t=1}^{T} \sum_i p_i^t \log r_i^t \geq \sum_{t=1}^{T} \log r_j^t - \varepsilon$$

$$\log\left(\frac{W_T(\text{new})}{W_0}\right) \geq \log \prod_{t=1}^{T} r_j^t - \varepsilon$$

$$\frac{W_T^{(\text{mwu})}}{W_0} \geq \prod_{t=1}^{T} r_j^t \, e^{-\varepsilon} \quad \forall \text{ stock } j \geq 1-\varepsilon$$

CRB