online learning with structured experts–a biased survey

Gábor Lugosi

ICREA and Pompeu Fabra University, Barcelona

on-line prediction

A repeated game between forecaster and environment.

At each round t,

- the forecaster chooses an action $I_t \in \{1, \ldots, N\}$;

  (actions are often called experts)
- the environment chooses losses $\ell_t(1), \ldots, \ell_t(N) \in [0, 1]$;
- the forecaster suffers loss $\ell_t(I_t)$.

The goal is to minimize the regret

$$R_n = \left(\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right) .$$

simplest example

Is it possible to make $(1/n)R_n \to 0$ for all loss assignments?

Let $N = 2$ and define, for all $t = 1, \ldots, n$,

$$\ell_t(1) = \begin{cases} 0 & \text{if } I_t = 2 \\ 1 & \text{if } I_t = 1 \end{cases}$$

and $\ell_t(2) = 1 - \ell_t(1)$.

Then

$$\sum_{t=1}^{n} \ell_t(I_t) = n \quad \text{and} \quad \min_{i=1,2} \sum_{t=1}^{n} \ell_t(i) \le \frac{n}{2}$$

so

$$\frac{1}{n} R_n \ge \frac{1}{2} \, .$$

randomized prediction

Key to solution: randomization.

At time t, the forecaster chooses a probability distribution
$\mathbf{p}_{t-1} = (p_{1,t-1}, \ldots, p_{N,t-1})$

and chooses action i with probability $p_{i,t-1}$.

Simplest model: all losses $\ell_s(i)$, $i = 1, \ldots, N$, $s < t$, are observed: full information.

Hannan and Blackwell

Hannan (1957) and Blackwell (1956) showed that the forecaster has a strategy such that

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t)-\min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right)\to 0$$

almost surely for all strategies of the environment.

basic ideas

expected loss of the forecaster:

$$\ell_t(p_{t-1}) = \sum_{i=1}^{N} p_{i,t-1}\ell_t(i) = \mathbb{E}_t\ell_t(I_t)$$

By martingale convergence,

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \sum_{t=1}^{n}\ell_t(p_{t-1})\right) = O_P(n^{-1/2})$$

so it suffices to study

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(p_{t-1}) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right)$$

Idea: assign a higher probability to better-performing actions. Vovk (1990), Littlestone and Warmuth (1989).

A popular choice is

$$p_{i,t-1} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(i)\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(k)\right)} \quad i = 1, \ldots, N .$$

where $\eta > 0$. Then

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)\right) = \sqrt{\frac{\ln N}{2n}}$$

with $\eta = \sqrt{8 \ln N / n}$.

proof

Let $L_{i,t} = \sum_{s=1}^{t} \ell_s(i)$ and

$$W_t = \sum_{i=1}^{N} w_{i,t} = \sum_{i=1}^{N} e^{-\eta L_{i,t}}$$

for $t \geq 1$, and $W_0 = N$. First observe that

$$\begin{aligned} \ln \frac{W_n}{W_0} &= \ln\left(\sum_{i=1}^{N} e^{-\eta L_{i,n}}\right) - \ln N \\ &\geq \ln\left(\max_{i=1,\ldots,N} e^{-\eta L_{i,n}}\right) - \ln N \\ &= -\eta \min_{i=1,\ldots,N} L_{i,n} - \ln N \, . \end{aligned}$$

On the other hand, for each $t = 1, \ldots, n$

$$\begin{aligned} \ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{i=1}^{N} w_{i,t-1} e^{-\eta \ell_t(i)}}{\sum_{j=1}^{N} w_{j,t-1}} \\ &\le -\eta \frac{\sum_{i=1}^{N} w_{i,t-1} \ell_t(i)}{\sum_{j=1}^{N} w_{j,t-1}} + \frac{\eta^2}{8} \\ &= -\eta \ell_t(p_{t-1}) + \frac{\eta^2}{8} \end{aligned}$$

by Hoeffding's inequality.

Hoeffding (1963): if $X \in [0, 1]$,

$$\ln \mathbb{E} e^{-\eta X} \le -\eta \mathbb{E} X + \frac{\eta^2}{8}$$

proof

for each $t = 1, \ldots, n$

$$\ln \frac{W_t}{W_{t-1}} \leq -\eta \ell_t(p_{t-1}) + \frac{\eta^2}{8}$$

Summing over $t = 1, \ldots, n$,

$$\ln \frac{W_n}{W_0} \leq -\eta \sum_{t=1}^{n} \ell_t(p_{t-1}) + \frac{\eta^2}{8} n \, .$$

Combining these, we get

$$\sum_{t=1}^{n} \ell_t(p_{t-1}) \leq \min_{i=1,\ldots,N} L_{i,n} + \frac{\ln N}{\eta} + \frac{\eta}{8} n$$

The upper bound is optimal: for all predictors,

$$\sup_{n,N,\ell_t(i)} \frac{\sum_{t=1}^n \ell_t(I_t) - \min_{i\le N} \sum_{t=1}^n \ell_t(i)}{\sqrt{(n/2)\ln N}} \ge 1 \ .$$

Idea: choose $\ell_t(i)$ to be i.i.d. symmetric Bernoulli coin flips.

$$\begin{aligned}
&\sup_{\ell_t(i)} \left(\sum_{t=1}^n \ell_t(I_t) - \min_{i\le N} \sum_{t=1}^n \ell_t(i) \right) \\
&\ge \ \mathbb{E}\left[\sum_{t=1}^n \ell_t(I_t) - \min_{i\le N} \sum_{t=1}^n \ell_t(i) \right] \\
&= \ \frac{n}{2} - \min_{i\le N} B_i
\end{aligned}$$

Where $B_1, \ldots, B_N$ are independent Binomial $(n, 1/2)$.
Use the central limit theorem.

follow the perturbed leader

$$I_t = \arg\min_{i=1,\dots,N} \sum_{s=1}^{t-1} \ell_s(i) + Z_{i,t}$$

where the $Z_{i,t}$ are random noise variables.

The original forecaster of Hannan (1957) is based on this idea.

If the $Z_{i,t}$ are i.i.d. uniform $[0, \sqrt{nN}]$, then

$$\frac{1}{n}R_n \leq 2\sqrt{\frac{N}{n}} + O_p(n^{-1/2}) \ .$$

If the $Z_{i,t}$ are i.i.d. with density $(\eta/2)e^{-\eta|z|}$, then for $\eta \approx \sqrt{\log N/n}$,

$$\frac{1}{n}R_n \leq c\sqrt{\frac{\log N}{n}} + O_p(n^{-1/2}) \ .$$

Kalai and Vempala (2003).

combinatorial experts

Often the class of experts is very large but has some combinatorial structure. Can the structure be exploited?

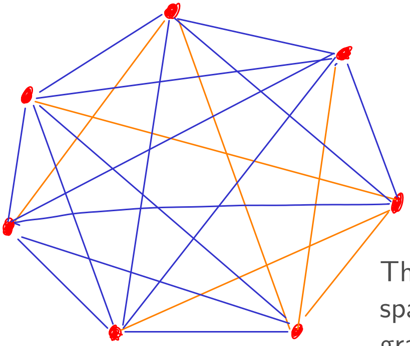
path planning. At each time instance, the forecaster chooses a path in a graph between two fixed nodes. Each edge has an associated loss. Loss of a path is the sum of the losses over the edges in the path.

N is huge!!!

Given a complete bipartite graph $K_{m,m}$, the forecaster chooses a perfect matching. The loss is the sum of the losses over the edges.

This image has been removed due to copyright restrictions.
Please see the image at
http://38.media.tumblr.com/tumblr_m0ol5tggjZ1qir7tc.gif

Helmbold and Warmuth (2007): full information case.



The forecaster chooses a spanning tree in the complete graph $\boldsymbol{K_m}$. The cost is the sum of the losses over the edges.

Formally, the class of experts is a set $\mathcal{S} \subset \{0,1\}^d$ of cardinality $|\mathcal{S}| = N$.

At each time t, a loss is assigned to each component: $\ell_t \in \mathbb{R}^d$.

Loss of expert $v \in \mathcal{S}$ is $\ell_t(v) = \ell_t^\top v$.

Forecaster chooses $I_t \in \mathcal{S}$.

The goal is to control the regret

$$\sum_{t=1}^{n} \ell_t(I_t) - \min_{k=1,\ldots,N} \sum_{t=1}^{n} \ell_t(k) \; .$$

One needs to draw a random element of $\mathcal{S}$ with distribution proportional to

$$w_t(v) = \exp\big(-\eta\, L_t(v)\big) = \exp\left(-\eta \sum_{s=1}^{t-1} \ell_t^\top v.\right)$$
$$= \prod_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{t,j} v_j\right) .$$

computing the exponentially weighted average forecaster

path planning: Sampling may be done by dynamic programming.

assignments: Sum of weights (partition function) is the permanent of a non-negative matrix. Sampling may be done by a FPAS of Jerrum, Sinclair, and Vigoda (2004).

spanning trees: Propp and Wilson (1998) define an exact sampling algorithm. Expected running time is the average hitting time of the Markov chain defined by the edge weights $w_t(v)$.

computing the follow-the-perturbed leader forecaster

In general, much easier. One only needs to solve a linear optimization problem over $\mathcal{S}$. This may be hard but it is well understood.

In our examples it becomes either a shortest path problem, or an assignment problem, or a minimum spanning tree problem.

follow the leader: random walk perturbation

Suppose N experts, no structure. Define

$$I_t = \arg\min_{i=1,\ldots,N} \sum_{s=1}^{t} (\ell_{i,s-1} + X_s)$$

where the X_s are either i.i.d. normal or ± 1 coinflips.

This is like follow-the-perturbed-leader but with random walk perturbation: $\sum_{s=1}^{t} X_t$.

Advantage: forecaster rarely changes actions!

If R_n is the regret and C_n is the number of times $I_t \neq I_{t-1}$, then

$$\mathbb{E}R_n \leq 2\mathbb{E}C_n \leq 8\sqrt{2n \log N} + 16 \log n + 16 \ .$$

Devroye, Lugosi, and Neu (2015).

Key tool: number of leader changes in N independent random walks with drift.

follow the leader: random walk perturbation

This also works in the "combinatorial" setting: just add an independent $\boldsymbol{N(0,d)}$ at each time to every component.

$$\mathbb{E}R_n = \widetilde{O}(B^{3/2}\sqrt{n \log d})$$

and

$$\mathbb{E}C_n = O(B\sqrt{n}\log d)\ ,$$

where $B = \max_{v \in \mathcal{S}} \|v\|_1$.

why exponentially weighted averages?

May be adapted to many different variants of the problem, including bandits, tracking, etc.

multi-armed bandits

The forecaster only observes $\ell_t(I_t)$ but not $\ell_t(i)$ for $i \neq I_t$.

Herbert Robbins (1952).

This image has been removed due to copyright restrictions. Please see the image at
https://en.wikipedia.org/wiki/Herbert_Robbins#/media/File:1966-HerbertRobbins.jpg

Trick: estimate $\ell_t(i)$ by

$$\widetilde{\ell}_t(i) = \frac{\ell_t(I_t)\mathbb{1}_{\{I_t=i\}}}{p_{I_t,t-1}}$$

This is an unbiased estimate:

$$\mathbb{E}_t\widetilde{\ell}_t(i) = \sum_{j=1}^{N} p_{j,t-1}\frac{\ell_t(j)\mathbb{1}_{\{j=i\}}}{p_{j,t-1}} = \ell_t(i)$$

Use the estimated losses to define exponential weights and mix with uniform (Auer, Cesa-Bianchi, Freund, and Schapire, 2002):

$$p_{i,t-1} = (1-\gamma)\underbrace{\frac{\exp\left(-\eta\sum_{s=1}^{t-1}\widetilde{\ell}_s(i)\right)}{\sum_{k=1}^{N}\exp\left(-\eta\sum_{s=1}^{t-1}\widetilde{\ell}_s(k)\right)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}}$$

$$\mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(p_{t-1}) - \min_{i\leq N}\sum_{t=1}^{n}\ell_t(i)\right) = O\left(\sqrt{\frac{N\ln N}{n}}\right) ,$$

Lower bound:

$$\sup_{\ell_t(i)} \mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)\right) \ge C\sqrt{\frac{N}{n}} \ ,$$

Dependence on N is not logarithmic anymore!

Audibert and Bubeck (2009) constructed a forecaster with

$$\max_{i \le N} \mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \sum_{t=1}^{n} \ell_t(i)\right) = O\left(\sqrt{\frac{N}{n}}\right) \ ,$$

calibration

Sequential probability assignment.

A binary sequence $x_1, x_2, \ldots$ is revealed one by one.

After observing $x_1, \ldots, x_{t-1}$, the forecaster issues prediction $I_t \in \{0, 1, \ldots, N\}$.

Meaning: "chance of rain is I_t/N".

Forecast is calibrated if

$$\left| \frac{\sum_{t=1}^n x_t \mathbb{1}_{\{I_t=i\}}}{\sum_{t=1}^n \mathbb{1}_{\{I_t=i\}}} - \frac{i}{N} \right| \le \frac{1}{2N} + o(1)$$

whenever $\limsup_n (1/n) \sum_{t=1}^n \mathbb{1}_{\{I_t=i\}} > 0$.

Is there a forecaster that is calibrated for all possible sequences? NO. (Dawid, 1985).

randomized calibration

However, if the forecaster is allowed to randomize then it is possible! (Foster and Vohra, 1997).

This can be achieved by a simple modification of any regret minimization procedure.

Set of actions (experts): $\{0, 1, \ldots, N\}$.

At time t, assign loss $\ell_t(i) = (x_t - i/N)^2$ to action i.

One can now define a forecaster. Minimizing regret is not sufficient.

internal regret

Recall that the (expected) regret is

$$\sum_{t=1}^{n} \ell_t(p_{t-1}) - \min_i \sum_{t=1}^{n} \ell_t(i) = \max_i \sum_{t=1}^{n} \sum_j p_{j,t} \left(\ell_t(j) - \ell_t(i)\right)$$

The internal regret is defined by

$$\max_{i,j} \sum_{t=1}^{n} p_{j,t} \left(\ell_t(j) - \ell_t(i)\right)$$

$$p_{j,t} \left(\ell_t(j) - \ell_t(i)\right) = \mathbb{E}_t \mathbb{1}_{\{I_t=j\}} \left(\ell_t(j) - \ell_t(i)\right)$$

is the expected regret of having taken action j instead of action i.

internal regret and calibration

By guaranteeing small internal regret, one obtains a calibrated forecaster.

This can be achieved by an exponentially weighted average forecaster defined over N^2 actions.

Can be extended even for calibration with checking rules.

For each round $t = 1, \ldots, n$,

- the environment chooses the next outcome $J_t \in \{1, \ldots, M\}$ without revealing it;
- the forecaster chooses a probability distribution p_{t-1} and draws an action $I_t \in \{1, \ldots, N\}$ according to p_{t-1};
- the forecaster incurs loss $\ell(I_t, J_t)$ and each action i incurs loss $\ell(i, J_t)$. None of these values is revealed to the forecaster;
- the feedback $h(I_t, J_t)$ is revealed to the forecaster.

$H = [h(i, j)]_{N \times M}$ is the feedback matrix.

$L = [\ell(i, j)]_{N \times M}$ is the loss matrix.

Dynamic pricing. Here $M = N$, and $L = [\ell(i,j)]_{N\times N}$ where

$$\ell(i,j) = \frac{(j-i)\mathbb{1}_{\{i\le j\}} + c\mathbb{1}_{\{i>j\}}}{N} .$$

and $h(i,j) = \mathbb{1}_{\{i>j\}}$ or

$$h(i,j) = a\mathbb{1}_{\{i\le j\}} + b\mathbb{1}_{\{i>j\}} , \quad i,j = 1,\ldots,N .$$

Multi-armed bandit problem. The only information the forecaster receives is his own loss: $H = L$.

Apple tasting. $N = M = 2$.

$$L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$H = \begin{bmatrix} a & a \\ b & c \end{bmatrix} .$$

The predictor only receives feedback when he chooses the second action.

Label efficient prediction. $N = 3$, $M = 2$.

$$L = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$H = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix} .$$

A forecaster first proposed by Piccolboni and Schindelhauer (2001).
Crucial assumption: $\boldsymbol{H}$ can be encoded such that there exists an $\boldsymbol{N \times N}$ matrix $\boldsymbol{K = [k(i,j)]_{N \times N}}$ such that

$$\boldsymbol{L = K \cdot H} .$$

Thus,

$$\boldsymbol{\ell(i,j) = \sum_{l=1}^{N} k(i,l) h(l,j)} .$$

Then we may estimate the losses by

$$\boldsymbol{\widetilde{\ell}(i,J_t) = \frac{k(i,I_t) h(I_t,J_t)}{p_{I_t,t}}} .$$

Observe

$$\mathbb{E}_t \widetilde{\ell}(i, J_t) = \sum_{k=1}^{N} p_{k,t-1} \frac{k(i,k)h(k,J_t)}{p_{k,t-1}}$$
$$= \sum_{k=1}^{N} k(i,k)h(k,J_t) = \ell(i,J_t) ,$$

$\widetilde{\ell}(i, J_t)$ is an unbiased estimate of $\ell(i, J_t)$.
Let

$$p_{i,t-1} = (1-\gamma)\frac{e^{-\eta \widetilde{L}_{i,t-1}}}{\sum_{k=1}^{N} e^{-\eta \widetilde{L}_{k,t-1}}} + \frac{\gamma}{N}$$

where $\widetilde{L}_{i,t} = \sum_{s=1}^{t} \widetilde{\ell}(i, J_t)$.

performance bound

With probability at least $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,J_t) - \min_{i=1,\ldots,N}\frac{1}{n}\sum_{t=1}^{n}\ell(i,J_t) \le Cn^{-1/3}N^{2/3}\sqrt{\ln(N/\delta)}\ .$$

where C depends on K. (Cesa-Bianchi, Lugosi, Stoltz (2006))

Hannan consistency is achieved with rate $O(n^{-1/3})$ whenever $L = K \cdot H$.

This solves the dynamic pricing problem.

Bartók, Pál, and Szepesvári (2010): if $M = 2$, only possible rates are $n^{-1/2}, n^{-1/3}, 1$

imperfect monitoring: a general framework

S is a finite set of signals.

Feedback matrix: $H : \{1, \ldots, N\} \times \{1, \ldots, M\} \to \mathcal{P}(S)$.

For each round $t = 1, 2 \ldots, n$,

- the environment chooses the next outcome $J_t \in \{1, \ldots, M\}$ without revealing it;
- the forecaster chooses p_{t-1} and draws an action $I_t \in \{1, \ldots, N\}$ according to it;
- the forecaster receives loss $\ell(I_t, J_t)$ and each action i suffers loss $\ell(i, J_t)$, none of these values is revealed to the forecaster;
- a feedback s_t drawn at random according to $H(I_t, J_t)$ is revealed to the forecaster.

Define

$$\ell(p, q) = \sum_{i,j} p_i q_j \ell(i, j)$$

$$H(\cdot, q) = (H(1, q), \ldots, H(N, q))$$

where $H(i, q) = \sum_j q_j H(i, j)$.

Denote by $\mathcal{F}$ the set of those Δ that can be written as $H(\cdot, q)$ for some q.

$\mathcal{F}$ is the set of "observable" vectors of signal distributions Δ.
The key quantity is

$$\rho(p, \Delta) = \max_{q \,:\, H(\cdot,q)=\Delta} \ell(p, q)$$

ρ is convex in p and concave in Δ.

The value of the base one-shot game is

$$\min_{p} \max_{q} \ell(p, q) = \min_{p} \max_{\Delta \in \mathcal{F}} \rho(p, \Delta)$$

If $\overline{q}_n$ is the empirical distribution of $J_1, \ldots, J_n$, even with the knowledge of $H(\cdot, \overline{q}_n)$ we cannot hope to do better than $\min_p \rho(p, H(\cdot, \overline{q}_n))$.

Rustichini (1999) proved that there exists a strategy such that for all strategies of the opponent, almost surely,

$$\limsup_{n \to \infty} \left(\frac{1}{n} \sum_{t=1,\ldots,n} \ell(I_t, J_t) - \min_{p} \rho\left(p, H(\cdot, \overline{q}_n)\right) \right) \leq 0$$

Rustichini's proof relies on an approachability theorem for a continuum of types (Mertens, Sorin, and Zamir, 1994).

It is non-constructive.

It does not imply any convergence rate.

Lugosi, Mannor, and Stoltz (2008) construct efficiently computable strategies that guarantee fast rates of convergence.

The class of actions is a set $\mathcal{S} \subset \{0,1\}^d$ of cardinality $|\mathcal{S}| = N$.

At each time t, a loss is assigned to each component: $\ell_t \in \mathbb{R}^d$.

Loss of expert $v \in \mathcal{S}$ is $\ell_t(v) = \ell_t^\top v$.

Forecaster chooses $I_t \in \mathcal{S}$.

The goal is to control the regret

$$\sum_{t=1}^{n} \ell_t(I_t) - \min_{k=1,\ldots,N} \sum_{t=1}^{n} \ell_t(k) .$$

Three models.

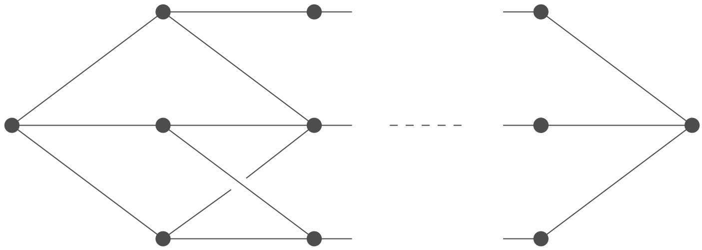
(Full information.) All d components of the loss vector are observed.

(Semi-bandit.) Only the components corresponding to the chosen object are observed.

(Bandit.) Only the total loss of the chosen object is observed.

Challenge: Is $O(n^{-1/2}\text{poly}(d))$ regret achievable for the semi-bandit and bandit problems?

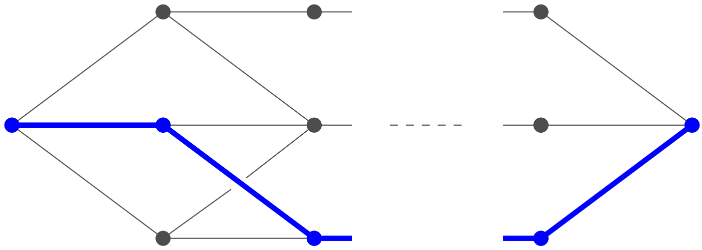
Adversary



Player

combinatorial prediction game

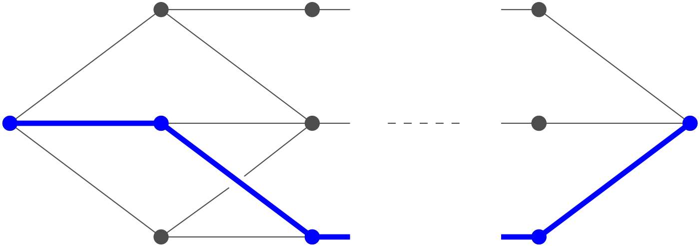
Adversary



Player ⟶

combinatorial prediction game

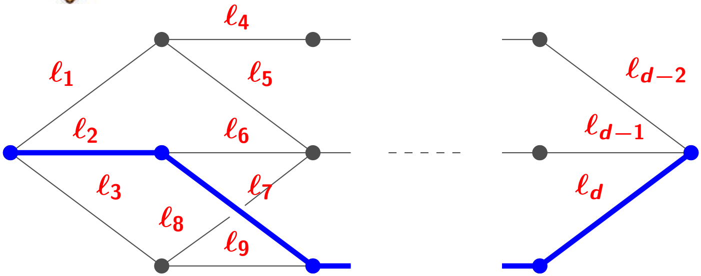
Adversary ⟶



Player ⟶

combinatorial prediction game

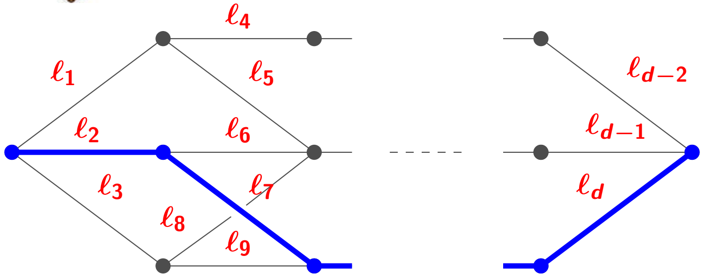
Adversary ⟶



Player ⟶

combinatorial prediction game

Adversary →



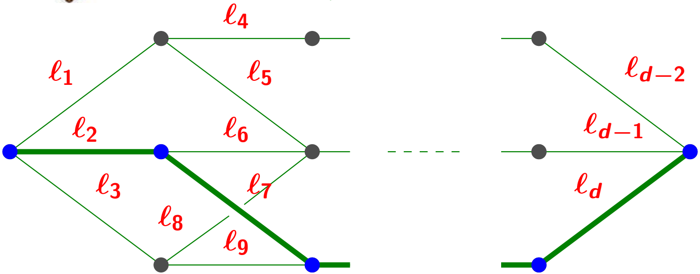
Player →

loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_d$

combinatorial prediction game

Adversary ⟶ 

Feedback: Full Info: $\ell_1, \ell_2, \ldots, \ell_d$



Player ⟶

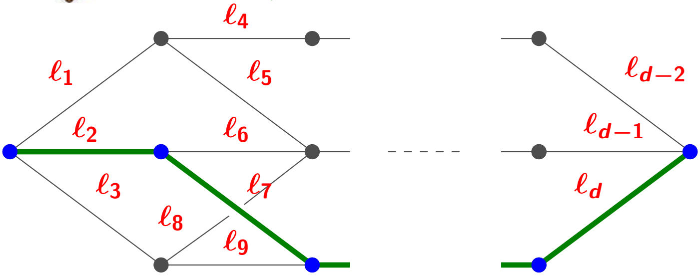
loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_d$

combinatorial prediction game

Adversary →

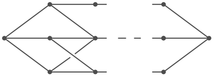
Feedback:

- Full Info: $\ell_1, \ell_2, \dots, \ell_d$
- Semi-Bandit: $\ell_2, \ell_7, \dots, \ell_d$
- Bandit: $\ell_2 + \ell_7 + \dots + \ell_d$

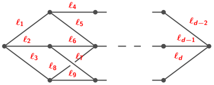


Player →

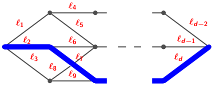
loss suffered: $\ell_2 + \ell_7 + \dots + \ell_d$



$\rightsquigarrow \quad \mathcal{S} \subset \{0,1\}^d$



$\rightsquigarrow \quad \ell_t \in \mathbb{R}_+^d$



$\rightsquigarrow \quad V_t \in \mathcal{S}$, loss suffered: $\ell_t^T V_t$

regret:

$$R_n = \mathbb{E} \sum_{t=1}^{n} \ell_t^T V_t - \min_{u \in S} \mathbb{E} \sum_{t=1}^{n} \ell_t^T u$$

loss assumption: $|\ell_t^T v| \leq 1$ for all $v \in \mathcal{S}$ and $t = 1, \ldots, n$.

At time t assign a weight $w_{t,i}$ to each $i = 1, \dots, d$.

The weight of each $v_k \in \mathcal{S}$ is

$$\overline{w}_t(k) = \prod_{i : v_k(i)=1} w_{t,i} \ .$$

Let $q_{t-1}(k) = \overline{w}_{t-1}(k) / \sum_{k=1}^{N} \overline{w}_{t-1}(k)$.

At each time t, draw K_t from the distribution

$$p_{t-1}(k) = (1-\gamma) q_{t-1}(k) + \gamma \mu(k)$$

where μ is a fixed distribution on $\mathcal{S}$ and $\gamma > 0$. Here

$$w_{t,i} = \exp\left(-\eta \widetilde{L}_{t,i}\right)$$

where $\widetilde{L}_{t,i} = \widetilde{\ell}_{1,i} + \cdots + \widetilde{\ell}_{t,i}$ and $\widetilde{\ell}_{t,i}$ is an estimated loss.

loss estimates

Dani, Hayes, and Kakade (2008).
Define the scaled incidence vector

$$\boldsymbol{X}_t = \ell_t(K_t) V_{K_t}$$

where K_t is distributed according to p_{t-1}.

Let $P_{t-1} = \mathbb{E}\big[V_{K_t} V_{K_t}^\top\big]$ be the $d \times d$ correlation matrix.
Hence

$$P_{t-1}(i,j) = \sum_{k : v_k(i) = v_k(j) = 1} p_{t-1}(k) \ .$$

Similarly, let Q_{t-1} and M be the correlation matrices of $\mathbb{E}\big[V V^\top\big]$ when V has law, q_{t-1} and μ. Then

$$P_{t-1}(i,j) = (1-\gamma) Q_{t-1}(i,j) + \gamma M(i,j) \ .$$

The vector of loss estimates is defined by

$$\widetilde{\ell}_t = P_{t-1}^{+} X_t$$

where P_{t-1}^{+} is the pseudo-inverse of P_{t-1}.

key properties

- $\boldsymbol{M}\,\boldsymbol{M}^{+}\boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{S}$.
- $\boldsymbol{Q}_{t-1}$ is positive semidefinite for every t.
- $\boldsymbol{P}_{t-1}\,\boldsymbol{P}_{t-1}^{+}\boldsymbol{v} = \boldsymbol{v}$ for all t and $\boldsymbol{v} \in \mathcal{S}$.

By definition,

$$\mathbb{E}_t\,\boldsymbol{X}_t = \boldsymbol{P}_{t-1}\,\boldsymbol{\ell}_t$$

and therefore

$$\mathbb{E}_t\,\widetilde{\boldsymbol{\ell}}_t = \boldsymbol{P}_{t-1}^{+}\mathbb{E}_t\,\boldsymbol{X}_t = \boldsymbol{\ell}_t$$

An unbiased estimate!

The regret of the forecaster satisfies

$$\frac{1}{n}\left(\mathbb{E}\widehat{L}_n - \min_{k=1,\ldots,N} L_n(k)\right) \le 2\sqrt{\left(\frac{2B^2}{d\lambda_{\min}(M)} + 1\right)\frac{d \ln N}{n}}\ .$$

where

$$\lambda_{\min}(M) = \min_{x \in \mathrm{span}(\mathcal{S}):\|x\|=1} x^T M x > 0$$

is the smallest "relevant" eigenvalue of M. (Cesa-Bianchi and Lugosi, 2009.)

Large $\lambda_{\min}(M)$ is needed to make sure no $|\widetilde{\ell}_{t,i}|$ is too large.

performance bound

Other bounds:

$B\sqrt{d \ln N / n}$ (Dani, Hayes, and Kakade). No condition on $\mathcal{S}$. Sampling is over a barycentric spanner.

$d\sqrt{(\theta \ln n)/n}$ (Abernethy, Hazan, and Rakhlin). Computationally efficient.

$$\lambda_{\min}(M) = \min_{x \in \mathrm{span}(\mathcal{S}) : \|x\| = 1} \mathbb{E}(V, x)^2 \ .$$

where V has distribution μ over $\mathcal{S}$.

In many cases it suffices to take μ uniform.

multitask bandit problem

The decision maker acts in m games in parallel.
In each game, the decision maker selects one of R possible actions.
After selecting the m actions, the sum of the losses is observed.

$$\lambda_{\min} = \frac{1}{R}$$

$$\max_k \mathbb{E}\left[\widehat{L}_n - L_n(k)\right] \leq 2m\sqrt{3nR\ln R}\ .$$

The price of only observing the sum of losses is a factor of m.

Generating a random joint action can be done in polynomial time.

Perfect matchings of $K_{m,m}$.

At each time one of the $N = m!$ perfect matchings of $K_{m,m}$ is selected.

$$\lambda_{\min}(M) = \frac{1}{m-1}$$

$$\max_k \mathbb{E}\left[\widehat{L}_n - L_n(k)\right] \leq 2m\sqrt{3n\ln(m!)} \; .$$

Only a factor of m worse than naive full-information bound.

In a network of m nodes, the cost of communication between two nodes joined by edge e is $\ell_t(e)$ at time t. At each time a minimal connected subnetwork (a spanning tree) is selected. The goal is to minimize the total cost. $N = m^{m-2}$.

$$\lambda_{\min}(M) = \frac{1}{m} - O\left(\frac{1}{m^2}\right) .$$

The entries of M are

$$\mathbb{P}\{V_i = 1\} = \frac{2}{m}$$

$$\mathbb{P}\left\{V_i = 1,\ V_j = 1\right\} = \frac{3}{m^2} \qquad \text{if } i \sim j$$

$$\mathbb{P}\left\{V_i = 1,\ V_j = 1\right\} = \frac{4}{m^2} \qquad \text{if } i \not\sim j .$$

At each time a central node of a network of m nodes is selected. Cost is the total cost of the edges adjacent to the node.

$$\lambda_{\min} \geq 1 - O\left(\frac{1}{m}\right) .$$

A balanced cut in K_{2m} is the collection of all edges between a set of m vertices and its complement. Each balanced cut has m^2 edges and there are $N = \binom{2m}{m}$ balanced cuts.

$$\lambda_{\min}(M) = \frac{1}{4} - O\left(\frac{1}{m^2}\right) .$$

Choosing from the exponentially weighted average distribution is equivalent to sampling from ferromagnetic Ising model. FPAS by Randall and Wilson (1999).

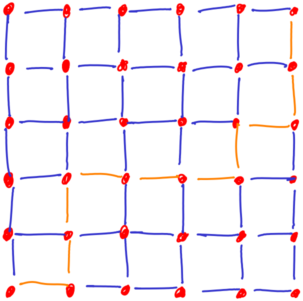
A Hamiltonian cycle in $\boldsymbol{K_m}$ is a cycle that visits each vertex exactly once and returns to the starting vertex. $\boldsymbol{N = (m-1)!}$

$$\lambda_{\min} \geq \frac{2}{m}$$

Efficient computation is hopeless.

In all these examples μ is uniform over $\mathcal{S}$.

For path planning it does not always work.



What is the optimal choice of μ?
What is the optimal way of exploration?

$$\overline{R}_n = \inf_{strategy} \max_{\mathcal{S} \subset \{0,1\}^d} \sup_{adversary} R_n$$

Theorem

Let $n \geq d^2$. In the full information and semi-bandit games, we have

$$0.008\, d\sqrt{n} \leq \overline{R}_n \leq d\sqrt{2n},$$

and in the bandit game,

$$0.01\, d^{3/2}\sqrt{n} \leq \overline{R}_n \leq 2\, d^{5/2}\sqrt{2n}.$$

upper bounds:

$\mathcal{D} = [0, +\infty)^d$, $F(x) = \frac{1}{\eta} \sum_{i=1}^{d} x_i \log x_i$ works for full information but it is only optimal up to a logarithmic factor in the semi-bandit case.
in the bandit case it does not work at all! Exponentially weighted average forecaster is used.

lower bounds:

careful construction of randomly chosen set $\mathcal{S}$ in each case.

Let $\mathcal{D}$ be a convex subset of $\mathbb{R}^d$ with nonempty interior $\mathrm{int}(\mathcal{D})$. A function $F : \mathcal{D} \to \mathbb{R}$ is Legendre if

- F is strictly convex and admits continuous first partial derivatives on $\mathrm{int}(\mathcal{D})$,
- For $u \in \partial\mathcal{D}$, and $v \in \mathrm{int}(\mathcal{D})$, we have

$$\lim_{s \to 0, s > 0} (u - v)^T \nabla F\big((1 - s)u + sv\big) = +\infty.$$

The Bregman divergence $D_F : \mathcal{D} \times \mathrm{int}(\mathcal{D})$ associated to a Legendre function F is

$$D_F(u, v) = F(u) - F(v) - (u - v)^T \nabla F(v).$$

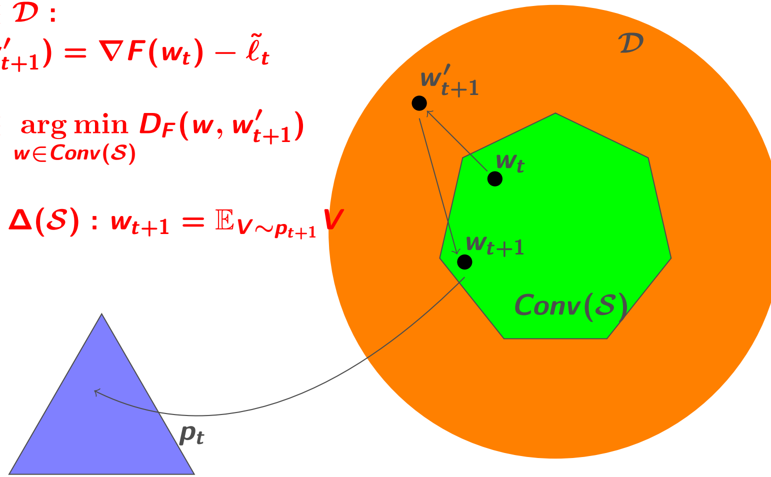
CLEB (Combinatorial LEarning with Bregman divergences)

Parameter: F Legendre on $\mathcal{D} \supset Conv(\mathcal{S})$

(1) $w'_{t+1} \in \mathcal{D}$:
$$\nabla F(w'_{t+1}) = \nabla F(w_t) - \tilde{\ell}_t$$

(2) $w_{t+1} \in \arg\min_{w \in Conv(\mathcal{S})} D_F(w, w'_{t+1})$

(3) $p_{t+1} \in \Delta(\mathcal{S})$: $w_{t+1} = \mathbb{E}_{V \sim p_{t+1}} V$



General regret bound for CLEB

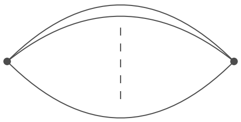
Theorem

If $\boldsymbol{F}$ *admits a Hessian* $\boldsymbol{\nabla^2 F}$ *always invertible then,*

$$\boldsymbol{R_n} \lessapprox \boldsymbol{diam_{D_F}(\mathcal{S})} + \mathbb{E}\sum_{t=1}^{n} \tilde{\ell}_t^T \left(\nabla^2 F(w_t)\right)^{-1} \tilde{\ell}_t.$$

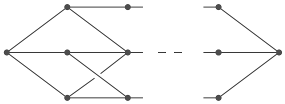
Different instances of CLEB: LinExp (Entropy Function)

$\mathcal{D} = [0, +\infty)^d$, $F(x) = \frac{1}{\eta} \sum_{i=1}^{d} x_i \log x_i$



Full Info: Exponentially weighted average

Semi-Bandit=Bandit: Exp3
Auer et al. [2002]



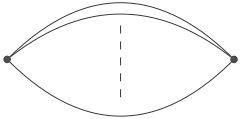
Full Info: Component Hedge
Koolen, Warmuth and Kivinen [2010]

Semi-Bandit: MW
Kale, Reyzin and Schapire [2010]

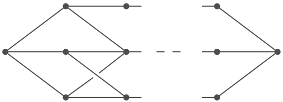
Bandit: new algorithm

Different instances of CLEB: LinINF (Exchangeable Hessian)

$\mathcal{D} = [0, +\infty)^d$, $F(x) = \sum_{i=1}^{d} \int_0^{x_i} \psi^{-1}(s) ds$



INF, Audibert and Bubeck [2009]



$$\begin{cases} \psi(x) = \exp(\eta x) : \text{LinExp} \\ \psi(x) = (-\eta x)^{-q}, q > 1 : \text{LinPoly} \end{cases}$$

Different instances of CLEB: Follow the regularized leader

$\mathcal{D} = Conv(\mathcal{S})$, then

$$w_{t+1} \in \arg\min_{w \in \mathcal{D}} \left(\sum_{s=1}^{t} \tilde{\ell}_s^T w + F(w) \right)$$

Particularly interesting choice: F self-concordant barrier function, Abernethy, Hazan and Rakhlin [2008]

18.657 Mathematics of Machine Learning
Fall 2015