Sequential learning – Lesson 1

Pierre Gaillard

January 14, 2022

INRIA

Practical information

Introduction

Full-information with linear loss functions

Application to prediction with expert advice

Teachers

Pierre Gaillard
INRIA Grenoble
pierre.gaillard@inria.fr

Remy Degenne
INRIA Lille
remy.degenne@inria.fr

Useful information

Webpage: `https://sequential-learning.github.io/`

MANDATORY REGISTRATION ON THE WEBPAGE

Validation method: 70% final exam, 30% homework.

Content of the class: mostly theoretical (algorithms and proofs), sequential learning with adversarial data, stochastic bandits, adversarial bandits

References

These monographs are available online. Links are on the webpage.

- Cesa-Bianchi and Lugosi, Prediction, learning, and games, 2006
- Shalev-Shwartz et al., "Online learning and online convex optimization", 2012
- Hazan et al., "Introduction to online convex optimization", 2016
- Lattimore and Szepesvári, "Bandit algorithms", 2019

Practical information

Introduction

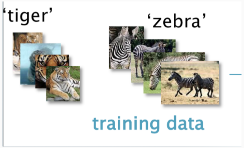
Full-information with linear loss functions

Application to prediction with expert advice

Classical Machine Learning

In classical supervised machine learning, the learner

1. observes training data with labels,
2. builds a program to minimize the training error
3. controls the error of new data if they are similar to the training data



$\longrightarrow$ Learning method $\longrightarrow$ Prediction on test data

Sequential Learning

In some applications, the environment may evolve over time and the data may be available sequentially.

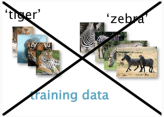
Spam detection: can be seen as a game between spammer and spam filters. Each trying to fool the other one. The data is possibly adversarial.

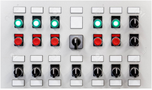
Necessity to take a robust approach by learning as ones goes along from experiences as more aspects of the problem are observed.

This is the goal of sequential learning (or sequential learning).

Sequential learning

In sequential learning, we do not have any training data.
Data are acquired and treated on the fly.
Feedbacks are received and algorithms updated step by step.



 $\Rightarrow$  $\Rightarrow$ zebra $\Rightarrow$ Change parameters $\Rightarrow$  $\Rightarrow$ ...

This field has received a lot of attention recently because of the possible applications coming from internet:

- ads to display,
- repeated auctions,
- spam detection,
- experts/algorithm aggregation

Setting of an online learning problem/online convex optimization

At each time step $t = 1, \ldots, T$

- the player observes a context $x_t \in \mathcal{X}$ (optional step)
- the player chooses an action $\theta_t \in \Theta$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [0, 1]$;
- the player suffers loss $\ell_t(\theta_t)$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\quad\to\quad$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t)$ $\quad\to\quad$ bandit feedback.

Goal. Minimize the cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

Example: Multi-armed bandit (Thomson, 2933)

A simple stochastic model:

- K arms (actions: here price signals)
- Each arm k is associated an unknown probability distribution with mean μ_k

μ_1 μ_2 μ_3 μ_4 μ_5

Setting: sequentially pick an arm k_t and get reward $X_{k_t,t}$ with mean μ_{k_t}

Goal: maximize the expected cumulative reward

$$\mathbb{E}\left[\sum_{t=1}^{T} X_{k_t,t}\right]$$

Exploration vs Exploitation trade-off.

Bandit applications

Maximize one's gains in casino? Hopeless . . .

Historical motivation (Thomson, 1933): clinical trials, for each patient t in a clinical study

- choose a treatment k_t
- observe response to the treatment $X_{k_t,t}$

Goal: maximize the number of patient healed (or find the best treatment)

Successful because of many applications coming from Internet: recommender systems, online advertisements,. . .

Setting of an online learning problem – Multi-armed bandits

At each time step $t = 1, \ldots, T$

- ~~the player observes a context $x_t \in \mathcal{X}$ (optional step)~~
- the player chooses an action $\theta_t = k_t \in \Theta := \{1, \ldots, K\}$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [0, 1]$ (by sampling the arms);
- the player suffers loss $\ell_t(\theta_t) = 1 - X_{k_t,t}$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\rightarrow$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t) = X_{k_t,t}$ $\rightarrow$ bandit feedback.

The goal of the player is to minimize his cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

Example 2: Prediction with expert advice

There is some sequence of observations $y_1, \dots, y_T \in [0,1]$ to be predicted step by step with the help of expert forecasts.

At each time step $t \geqslant 1$

- the environment reveals experts forecasts $x_t(k)$ for $k = 1, \dots, K$
- the player chooses a weight vector $p_t \in \Delta_K \stackrel{\text{def}}{=} \{p \in [0,1]^K : \sum_{k=1}^K p_k = 1\}$ (here θ_t is denoted p_t and $\Theta = \Delta_K$)
- the player forecasts $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$
- the environment reveals $y_t \in [0,1]$ and the player suffers loss $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ where $\ell : [0,1]^2 \to [0,1]$ is a loss function.

Considering $\Theta := \Delta_K$ and $\theta_t := p_t$, we recover the general setting. The inputs correspond to the expert advice $x_t(k)$ that are often revealed before the learner makes his decision p_t.

Example 2: Prediction with expert advice

There is some sequence of observations $y_1, \dots, y_T \in [0,1]$ to be predicted step by step with the help of expert forecasts.

At each time step $t \geqslant 1$

- the environment reveals experts forecasts $x_t(k)$ for $k = 1, \dots, K$
- the player chooses a weight vector $p_t \in \Delta_K \stackrel{\text{def}}{=} \{p \in [0,1]^K : \sum_{k=1}^K p_k = 1\}$ (here θ_t is denoted p_t and $\Theta = \Delta_K$)
- the player forecasts $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$
- the environment reveals $y_t \in [0,1]$ and the player suffers loss $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ where $\ell : [0,1]^2 \to [0,1]$ is a loss function.

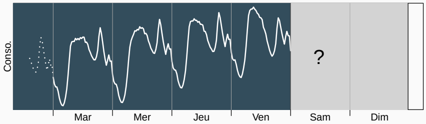
Player's performance is then measured via a loss function $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ which measures the distance between the prediction $\widehat{y}_t$ and the output y_t:

- squared loss $\ell(\widehat{y}_t, y_t) = (\widehat{y}_t - y_t)^2$
- absolute loss $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t|$
- absolute percentage of error $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t| / |y_t|$
- pinball loss.

All these loss functions are convex, which will play an important role in the analysis.

Example: Prediction with expert advice for electricity forecasting

Short term prediction (one day ahead) of the French electricity consumption



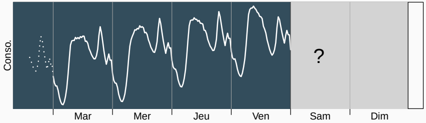
Important because electricity is hard to store.

Production

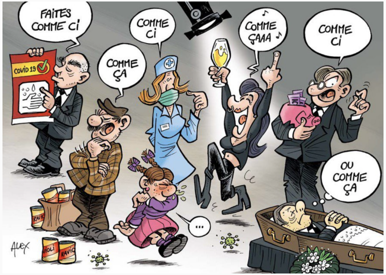
Demand

Example: Prediction with expert advice for electricity forecasting

Short term prediction (one day ahead) of the French electricity consumption



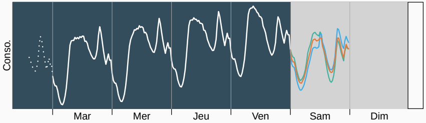
Many experts (statisticians or data scientists) design prediction models:



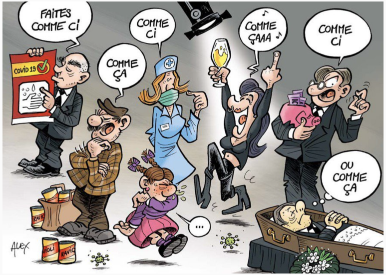
Simultaneously, the French electricity market is evolving (electric cars,. . .)

Example: Prediction with expert advice for electricity forecasting

Short term prediction (one day ahead) of the French electricity consumption



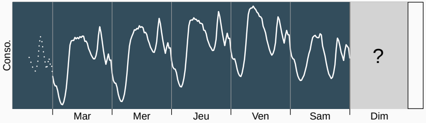
Many experts (statisticians or data scientists) design prediction models:



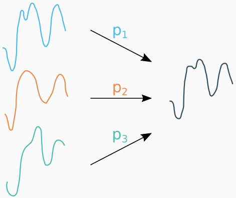
Simultaneously, the French electricity market is evolving (electric cars,. . .)

Example: Prediction with expert advice for electricity forecasting

Short term prediction (one day ahead) of the French electricity consumption



Combine the predictions using adaptive methods:



Each day,

1. Assign a weight to each expert based on past performance

$$\theta_t = \text{weight vector}$$

2. Predict the weighted average $\widehat{y}_t = \langle \theta_t, x_t \rangle$ and suffer loss

$$\ell_t(\theta_t) = \big(y_t - \widehat{y}_t\big)^2$$

How to measure the performance? The regret

If the environment chooses large losses $\ell_t(x)$ for all decisions $\theta \in \Theta$, it is impossible for the player to ensure small cumulative loss.

$\to$ Relative criterion: the regret of the player is the difference between the cumulative loss he incurred and that of the best fixed decision in hindsight.

Definition (Regret)

The regret of the player with respect to a fixed parameter $\theta^* \in \Theta$ after T time steps is

$$R_T(\theta^*) \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(\theta^*) \, .$$

The regret (or uniform regret) is defined as $R_T \stackrel{\text{def}}{=} \sup_{\theta^* \in \Theta} R_T(\theta^*)$.

Regret decomposition

We have some approximation-estimation decomposition:

$$\sum_{t=1}^{T} \ell_t(\theta_t) = \underbrace{\inf_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)}_{\text{Approximation error = how good the possible actions are.}} + \underbrace{R_T}_{\text{Sequential estimation error of the best action}}$$

We will focus on the regret in these lectures.

The goal of the player is to ensure a sublinear regret $R_T = o(T)$ as $T \to \infty$ and this for any possible sequence of losses $\ell_1, \ldots, \ell_T$.
$\to$ the average performance of the player will approach on the long term the one of the best decision.

Adversarial / Stochastic setting

The losses ℓ_t are unknown to the player beforehand and may be:

- Adversarial setting (lessons 1, 2, and 3): No stochastic assumption on the process generating the losses ℓ_t. The latter are deterministic and may be chosen by some adversary. Typically, the problem can be seen as a game between the player who aims at optimizing with respect to $\theta_1, \dots, \theta_T$ against an environment who aims at mazimizing with respect to $loss_t, \dots, loss_T$ and θ^*. Players's goal is to control the quantity:

$$\inf_{\theta_1} \sup_{\ell_1} \inf_{\theta_2} \sup_{\ell_2} \dots \inf_{\theta_T} \sup_{\ell_T} \sup_{\theta^* \in \Theta} R_T(\theta^*) .$$

- Stochastic setting (lessons 4, 5, and 6): the losses are generated by some stochastic process (e.g., i.i.d.). The regret bounds hold then in expectation or with high probability.

Why a different loss at every round t?

This may be caused by many phenomena, e.g. by

- some observation to be predicted if $\ell_t(x) = \ell(x, y_t)$. For instance, if the goal is to predict the evolution of the temperature $y_1, \ldots, y_T$, the latter changes over time and a prediction x is evaluated with $\ell_t(x) = (x - y_t)^2$.
- noise: the environment is stochastic and the variation over time t models some noise effect.
- a changing environment. For instance, if the player is playing a game against some adversary that evolves and adapts to its strategy. A typical example is the case of spam detections. If the player tries to detect spams, while some spammers (the environment) try at the same time to fool the player with new spam strategies.

Exercise: what about best θ_t^* at every round?

Regret

$$R_T = \sum_{t=1}^{T} \ell_t(\theta_t) - \inf_{\theta^* \in \Theta} \sum_{t=1}^{T} \ell_t(\theta^*)$$

Instead considering the regret with respect to a fixed $\theta^* \in \Theta$, one would be tempted to minimize the quantity

$$R_T^* \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \inf_{\theta \in \Theta} \ell_t(\theta)$$

where the infimum is inside the sum.

Exercise: Show that the environment can ensure R_T^* to be linear in T by choosing properly the loss functions ℓ_t.

Practical information

Introduction

Full-information with linear loss functions

Application to prediction with expert advice

Full-information with linear loss functions

We will start with the simple case where the decision set Θ is the K-dimensional simplex

$$\Delta_K \stackrel{\text{def}}{=} \left\{ p \in [0,1]^K : \sum_{k=1}^{K} p_k = 1 \right\}. \qquad \text{(decision set)}$$

Since the decisions θ_t are probability distributions in $\Theta = \Delta_K$, in this part we will denote them by p_t instead of θ_t. We assume that the loss functions ℓ_t are linear

$$\forall p \in \Theta, \qquad \ell_t(p) = \sum_{k=1}^{K} p(k) g_t(k) \in [-1, 1] \qquad \text{(linear loss)}$$

where $g_t = (g_t(1), \ldots, g_t(K)) \in [-1,1]^K$ is a loss vector chosen by the environment at round t.

How to choose the weights

At round t the player needs to choose a weight vector $p_t \in \Delta_K$.

How to choose the weights? The player should

- give more weight to actions that performed well in the past.
- not give all the weight to the current best action, otherwise it would not work (see Exercise next).

The exponentially weighted average forecaster (EWA) also called Hedge performs this trade-off by choosing a weight that decreases exponentially fast with the past errors.

The exponentially weighted average forecaster (EWA)

The exponentially weighted average forecaster

Parameter: $\eta > 0$

Initialize: $p_1 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$

For $t = 1, \dots, T$

- select p_t; incur loss $\ell_t(p_t) = p_t^\top g_t$ and observe $g_t \in [-1, 1]^K$;
- update for all $k \in \{1, \dots, K\}$

$$p_{t+1}(k) = \frac{e^{-\eta \sum_{s=1}^{t} g_s(k)}}{\sum_{j=1}^{K} e^{-\eta \sum_{s=1}^{t} g_s(j)}} .$$

Exercise

Consider the strategy, called “Follow The Leader” (FTL) that puts all the mass on the best action so far:

$$p_t \in \arg\min_{p \in \Theta} \sum_{s=1}^{t-1} \ell_s(p) \,. \tag{FTL}$$

Exercise:

1. Show that $p_t(k) > 0$ implies that $k \in \arg\min_j \sum_{s=1}^{t-1} g_s(j)$
2. Show that the regret of FTL might be linear: i.e., there exists a sequence $g_1, \dots, g_T \in [-1,1]^K$ such that $R_T \geqslant \Omega(T)$.

You have 15 min to think about this exercise

Solution

Consider the strategy, called "Follow The Leader" (FTL) that puts all the mass on the best action so far:

$$p_t \in \arg\min_{p\in\Theta} \sum_{s=1}^{t-1} \ell_s(p)\,. \tag{FTL}$$

Exercise:

1. Show that $p_t(k) > 0$ implies that $k \in \arg\min_j \sum_{s=1}^{t-1} g_s(j)$

Solution

Assume that there exists $k \in [K]$ such that $p_t(k) > 0$ and $k \notin \arg\min_j \sum_{s=1}^{t-1} g_s(j)$. Then, there exists $k' \neq k$ such that $\sum_{s=1}^{t-1} g_s(k') < \sum_{s=1}^{t-1} g_s(k)$. Therefore,

$$\begin{aligned}\sum_{s=1}^{t-1} \ell_s(p_t) &= \sum_{s=1}^{t-1}\sum_{j=1}^{K} p_s(j)g_s(j) = \sum_{s=1}^{t-1}\sum_{j\neq k} p_s(j)g_s(j) + p_s(k)\sum_{s=1}^{t-1} g_s(k)\\ &> \sum_{s=1}^{t-1}\sum_{j\neq k} p_s(j)g_s(j) + p_s(k')\sum_{s=1}^{t-1} g_s(k) = \sum_{s=1}^{t-1} \ell_s(q_t)\,,\end{aligned}$$

where $q_t(j) = p_t(j)$ if $j \notin \{k, k'\}$ and $q_t(k) = 0$ and $q_t(k') = p_t(k') + q_t(k')$. This yields a contradiction.

Solution

Consider the strategy, called "Follow The Leader" (FTL) that puts all the mass on the best action so far:

$$p_t \in \arg\min_{p\in\Theta} \sum_{s=1}^{t-1} \ell_s(p)\,. \tag{FTL}$$

Exercise:

1. Show that $p_t(k) > 0$ implies that $k \in \arg\min_j \sum_{s=1}^{t-1} g_s(j)$
2. Show that the regret of FTL might be linear: i.e., there exists a sequence $g_1, \dots, g_T \in [0,1]^K$ such that $R_T \geqslant \Omega(T)$.

Solution

It suffices to choose $g_t(k) = 1$ if $p_t(k) > 0$ and $g_t(k) = 0$ otherwise. The cumulative loss of FTL is T while there exists an action with cumulative loss smaller then T/K.

Regret guarantee for EWA

Theorem 1 (Regret bound for EWA)

Let $T \geqslant 1$. For all sequences of loss vectors $g_1, \dots, g_T \in [-1, 1]^K$, EWA achieves the bound

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^{T} \ell_t(p) \leqslant \eta \sum_{t=1}^{T} \sum_{k=1}^{K} p_t(k) g_t(k)^2 + \frac{\log K}{\eta}, \tag{1}$$

where we recall $\ell_t : p \in \Delta_K \mapsto p^\top g_t$.

Therefore, for the choice $\eta = \sqrt{\frac{\log K}{T}}$, EWA satisfies the regret bound $R_T \leqslant 2\sqrt{T \log K}$.

This regret bound is optimal (see [1]).

Exercise: Generalize the above theorem when the losses $g_1, \dots, g_T \in [-B, B]^K$ for some $B > 0$.

[1] Cesa-Bianchi and Lugosi, Prediction, learning, and games, 2006.

Proof (Step 1 - Reformulation of the regret for linear losses)

First, we remark that by definition of $\ell_t : p \mapsto p \cdot g_t$ we have

$$\begin{aligned} R_T &\stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^{T} \ell_t(p) \\ &= \sum_{t=1}^{T} p_t \cdot g_t - \min_{p \in \Delta_K} \sum_{t=1}^{T} p \cdot g_t \\ &= \sum_{t=1}^{T} p_t \cdot g_t - \min_{p \in \Delta_K} \sum_{k=1}^{K} \sum_{t=1}^{T} p(k) g_t(k) \,. \end{aligned}$$

Now, we can see that the minimum over $p \in \Delta_K$ is reached on a corner of the simplex. Therefore

$$R_T = \sum_{t=1}^{T} p_t \cdot g_t - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} g_t(k) \,.$$

Proof (Step 2 – Upper-bound of W_T)

We denote $W_t(j) = e^{-\eta \sum_{s=1}^t g_t(j)}$ and $W_t = \sum_{j=1}^K W_t(j)$. The proof will consist in upper-bounding and lower-bounding W_T. We have

$$
\begin{aligned}
W_t &= \sum_{j=1}^K W_{t-1}(j) e^{-\eta g_t(j)} && \leftarrow \quad W_t^{(j)} = W_{t-1}(j) e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^K \frac{W_{t-1}(j)}{W_{t-1}} e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^K p_t(j) e^{-\eta g_t(j)} && \leftarrow \quad p_t(j) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(j)}}{\sum_{k=1}^K e^{-\eta \sum_{s=1}^{t-1} g_s(k)}} = \frac{W_{t-1}(j)}{W_{t-1}} \\
&\leqslant W_{t-1} \sum_{j=1}^K p_t(j) \big(1 - \eta g_t(j) + \eta^2 g_t(j)^2\big) && \leftarrow \quad e^x \leqslant 1 + x + x^2 \text{ for } x \leqslant 1 \\
&= W_{t-1} (1 - \eta p_t \cdot g_t + \eta^2 p_t \cdot g_t^2)\,,
\end{aligned}
$$

where we assumed in the inequality $-\eta g_t(j) \leqslant 1$ and where we denote $g_t = (g_t(1), \dots, g_t(K))$, $g_t^2 = \big(g_t(1)^2, \dots, g_t(K)^2\big)$ and $p_t = (p_t(1), \dots, p_t(K))$.

Proof (Step 2 - Upper-bound of W_T)

Now, using $1 + x \leqslant e^x$, we get:

$$W_t \leqslant W_{t-1}(1 - \eta p_t \cdot g_t + \eta^2 p_t \cdot g_t^2) \leqslant W_{t-1} \exp\left(-\eta p_t \cdot g_t + \eta^2 p_t \cdot g_t^2\right).$$

By induction on $t = 1, \ldots, T$, this yields using $W_0 = K$

$$W_T \leqslant K \exp\left(-\eta \sum_{t=1}^{T} p_t \cdot g_t + \eta^2 \sum_{t=1}^{T} p_t \cdot g_t^2\right). \tag{2}$$

Proof (Step 3 – Lower-bound of W_T)

On the other hand, upper-bounding the maximum with the sum,

$$\exp\left(-\eta \min_{j\in[K]} \sum_{t=1}^{T} g_t(j)\right) \leqslant \sum_{j=1}^{K} \exp\left(-\eta \sum_{t=1}^{T} g_t(j)\right) \leqslant W_T\,.$$

Combining the above inequality with Inequality (2) and taking the log, we get

$$-\eta \min_{j\in[K]} \sum_{t=1}^{T} g_t(j) \leqslant -\eta \sum_{t=1}^{T} p_t \cdot g_t + \eta^2 \sum_{t=1}^{T} p_t \cdot g_t^2 + \log K\,. \tag{3}$$

Dividing by η and reorganizing the terms proves the first inequality:

$$R_T = \sum_{t=1}^{T} p_t \cdot g_t - \min_{1\leqslant j\leqslant K} \sum_{t=1}^{T} g_t(j) \leqslant \eta \sum_{t=1}^{T} p_t \cdot g_t^2 + \frac{\log K}{\eta}$$

Optimizing η and upper-bounding $p_t \cdot g_t^2 \leqslant 1$ concludes the second inequality. $\square$

Regret guarantee for EWA

Theorem 1 (Regret bound for EWA)

Let $T \geqslant 1$. For all sequences of loss vectors $g_1, \dots, g_T \in [-1,1]^K$, EWA achieves the bound

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^{T} \ell_t(p) \leqslant \eta \sum_{t=1}^{T} \sum_{k=1}^{K} p_t(k) g_t(k)^2 + \frac{\log K}{\eta}, \tag{1}$$

where we recall $\ell_t : p \in \Delta_K \mapsto p^\top g_t$.

Therefore, for the choice $\eta = \sqrt{\frac{\log K}{T}}$, EWA satisfies the regret bound $R_T \leqslant 2\sqrt{T \log K}$.

This regret bound is optimal (see [1]).

Exercise: Generalize the above theorem when the losses $g_1, \dots, g_T \in [-B, B]^K$ for some $B > 0$.

[1] Cesa-Bianchi and Lugosi, Prediction, learning, and games, 2006.

Anytime algorithm

The previous algorithms EWA depends on a parameter $\eta > 0$ that needs to be optimized according to K and T. For instance, for EWA using the value

$$\eta = \sqrt{\frac{\log K}{KT}}\,.$$

The bound of Theorem 1 is only valid for horizon T.

However, the learner might not know the time horizon in advance and one might want an algorithm with guarantees valid simultaneously for all $T \geqslant 1$.

We can avoid the assumption that T is known in advance, at the cost of a constant factor, by using the so-called doubling trick.

Anytime algorithm: the doubling trick

Whenever we reach a time step t which is a power of 2, we restart the algorithm (forgetting all the information gained in the past) setting η to $\sqrt{\log K/t}$. Let us denote EWA-doubling this algorithm.

Theorem 2 (Anytime bound on the regret)

For all $T \geqslant 1$, the regret of EWA-doubling is then upper-bounded as:

$$R_T \leqslant 7\sqrt{T \log K}\,.$$

The same trick can be used to turn most online algorithms into anytime algorithms (even in more general settings: bandits, general loss,...).

We can use the doubling trick whenever we have an algorithm with a regret of order $\mathcal{O}(T^\alpha)$ for some $\alpha > 0$ with a known horizon T to turn it into an algorithm with a regret $\mathcal{O}(T^\alpha)$ for all $T \geqslant 1$.

Proof

For simplicity we assume $T = 2^{M+1} - 1$.

You have 15 min to think about the proof

Proof

For simplicity we assume $T = 2^{M+1} - 1$. The regret of EWA-doubling is then upper-bounded as:

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^{T} \ell_t(p) \\ &\leqslant \sum_{t=1}^{T} \ell_t(p_t) - \sum_{m=0}^{M} \min_{p \in \Delta_K} \sum_{t=2^m}^{2^{m+1}-1} \ell_t(p) \\ &= \sum_{m=0}^{M} \underbrace{\sum_{t=2^m}^{2^{m+1}-1} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=2^m}^{2^{m+1}-1} \ell_t(p)}_{R_m} . \end{aligned}$$

Now, we remark that each term R_m corresponds to the expected regret of an instance of EWA over the 2^m rounds $t = 2^m, \dots, 2^{m+1} - 1$ and run with the optimal parameter $\eta = \sqrt{\log K / 2^m}$. Therefore, using Theorem 1, we get $R_m \leqslant 2\sqrt{2^m \log K}$, which yields:

$$R_T \leqslant \sum_{m=0}^{M} 2\sqrt{2^m \log K} \leqslant 2(1 + \sqrt{2})\sqrt{2^{M+1} \log K} \leqslant 7\sqrt{T \log K} .$$

Anytime algorithm: time-varying parameter

Another solution is to use time-varying parameters η_t replacing T with the current value of t. The analysis is however less straightforward.

Exercise: Prove a regret bound for the time-varying choice $\eta_t = \sqrt{\log K/t}$ in EWA.

Improvement for small losses

Theorem 3 (Improvement for small losses)

For well-chosen learning rate $\eta \approx \sqrt{(\log K)/L_T^*}$, *if the losses take values in* $[0,1]$, *EWA achieves*

$$R_T \lesssim \sqrt{(\log K) L_T^*}\,.$$

The regret is small whenever some decision achieves a small cumulative loss.

Proof

Let's define $\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^T \ell_t(p_t)$ the loss of the algorithm and $L_T^* \stackrel{\text{def}}{=} \min_{p \in \Delta_K} \sum_{t=1}^T \ell_t(p)$. Starting from the first inequality in Theorem 1,

$$\begin{aligned} R_T \stackrel{\text{def}}{=} \widehat{L}_T - L_T^* \quad & \leqslant \quad \frac{\log K}{\eta} + \eta \sum_{t=1}^T p_t \cdot g_t^2 \,, \\ & \leqslant \quad \frac{\log K}{\eta} + \eta \sum_{t=1}^T p_t \cdot g_t = \frac{\log K}{\eta} + \eta \widehat{L}_T \,. \end{aligned}$$

Therefore, rearranging the terms

$$(1-\eta)\widehat{L}_T - (1-\eta)L_T^* \leqslant \frac{\log K}{\eta} + \eta L_T^* \,,$$

which implies

$$R_T \leqslant \frac{\log K}{\eta(1-\eta)} + \frac{\eta}{1-\eta} L_T^* \,.$$

Optimising in $\eta \approx \sqrt{(\log K)/L_T^*}$ concludes. $\square$

Practical information

Introduction

Full-information with linear loss functions

Application to prediction with expert advice

Reminder of the setting of prediction with expert advice

At each time step $t \geqslant 1$

- the environment reveals experts forecasts $x_t(k)$ for $k = 1, \dots, K$
- the player chooses a weight vector $p_t \in \Delta_K \stackrel{\text{def}}{=} \{p \in [0,1]^K : \sum_{k=1}^K p_k = 1\}$ (here θ_t is denoted p_t and $\Theta = \Delta_K$)
- the player forecasts $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$
- the environment reveals $y_t \in [0,1]$ and the player suffers loss $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ where $\ell : [0,1]^2 \to [0,1]$ is a loss function.

The goal is to minimize the regret with respect to the best expert

$$R_T^{\text{expert}} \stackrel{\text{def}}{=} \sum_{t=1}^T \ell(\widehat{y}_t, y_t) - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^T \ell(x_t(k), y_t),$$

where $\widehat{y}_t = p_t \cdot x_t$ are the prediction of the algorithm and y_t the observations to be predicted sequentially.

Reminder of the setting of prediction with expert advice

At each time step $t \geqslant 1$

- the environment reveals experts forecasts $x_t(k)$ for $k = 1, \dots, K$
- the player chooses a weight vector $p_t \in \Delta_K \stackrel{\text{def}}{=} \{p \in [0,1]^K : \sum_{k=1}^K p_k = 1\}$ (here θ_t is denoted p_t and $\Theta = \Delta_K$)
- the player forecasts $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$
- the environment reveals $y_t \in [0,1]$ and the player suffers loss $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ where $\ell : [0,1]^2 \to [0,1]$ is a loss function.

Player's performance is then measured via a loss function $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ which measures the distance between the prediction $\widehat{y}_t$ and the output y_t:

- squared loss $\ell(\widehat{y}_t, y_t) = (\widehat{y}_t - y_t)^2$
- absolute loss $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t|$
- absolute percentage of error $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t| / |y_t|$
- pinball loss.

All these loss functions are convex, how can we apply our analysis for linear losses?

Prediction with expert advice with convex loss function ℓ.

We state bellow a corrolary to Theorem 1 when the loss functions $\ell(\cdot,\cdot)$ are convex in there first argument.

Corollary 1 (Regret of EWA for prediction with expert advice and convex loss)

Let $T \geqslant 1$. Assume that the loss function $\ell : (x, y) \in \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ is convex and takes values in $[-1, 1]$. Then, EWA applied with the vector vectors $g_t = \big(\ell(x_t(1), y_t), \dots, \ell(x_t(K), y_t)\big) \in [-1, 1]^K$ has a regret upper-bounded by

$$R_T^{\text{expert}} \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} \ell(x_t(k), y_t) \leqslant 2\sqrt{T \log K}$$

where $\widehat{y}_t = p_t \cdot x_t$ and were $\eta > 0$ is well-tuned.

Therefore, the average error of the algorithm will converge to the average error of the best expert. This is the case for the square loss, the absolute loss or the absolute percentage of error.

Proof

It suffices to remark that by convexity of $\ell(\cdot, \cdot)$ in its first argument

$$\begin{aligned} R_T^{\text{expert}} &= \sum_{t=1}^{T} \ell(p_t \cdot x_t, y_t) - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} \ell(x_t(k), y_t) \\ &\leqslant \sum_{t=1}^{T} p_t \cdot g_t - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} g_t(k) \stackrel{\text{def}}{=} R_T . \end{aligned}$$

The result is then obtained by Theorem 1. $\square$

Exp-concavity

Definition (η-exp-concavity)

For $\eta \in \mathbb{R}$, a function f is said to be η-exp-concave if $x \mapsto e^{-\eta f(x)}$ is concave.

Properties:

- Exp-concavity $\Rightarrow$ convexity because $-\log$ is convex and decreasing.
- Strong convexity $+$ bounded domain and gradients $\Rightarrow$ exp-concavity.
- η-exp-concavity $\Rightarrow$ η'-exp-concavity for $0 \leqslant \eta' \leqslant \eta$.

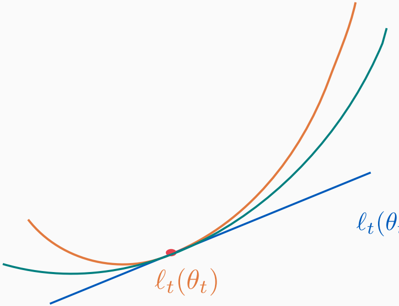
Many losses are exp-concave:

- strongly convex losses on bounded domain
- squared loss is $\frac{1}{2Y^2}$-expconcave on $[0, Y]$
- logistic loss
- relative entropy

Exercise: Prove the above the above facts.

Exercise: Exp-concavity implies a quadratic lower-bound

Show that if the loss ℓ_t is η-exp-concave with gradients bounded by G and a domain diameter bounded by D, then it can be lower-bounded by a quadratic approximation: for all $\theta, \theta' \in \Theta$, and $\gamma \leqslant \frac{1}{2}\min\left\{\eta, \frac{1}{4GD}\right\}$



$$\ell_t(\theta_t)$$

$$\geq$$

$$\ell_t(\theta_t) + \nabla \ell_t(\theta_t)^\top(\theta - \theta_t) + \gamma\big(\nabla \ell_t(\theta_t)^\top(\theta - \theta_t)\big)^2$$

$$\ell_t(\theta_t) + \nabla \ell_t(\theta_t)^\top(\theta - \theta_t)$$

Constant regret for exp-concave loss functions

Corollary 2 (Regret of EWA for prediction with expert advice and exp-concave loss)

In the setting of prediction with expert advice, if the loss functions $\ell(\cdot, y_t)$ are η-exp-concave for all y_t, then EWA run with vectors $g_t = \big(\ell(x_t(1), y_t), \ldots, \ell(x_t(K), y_t)\big) \in \mathbb{R}^K$ with parameter $\eta > 0$ satisfies

$$R_T^{\text{expert}} \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) - \min_{1 \leqslant k \leqslant K} \sum_{t=1}^{T} \ell(x_t(k), y_t) \leqslant \frac{\log K}{\eta},$$

for all $T \geqslant 1$.

The worst-case regret does not increase with T but grows logarithmically in the dimension K.

Proof (Step 1)

We define $W_t(i) = e^{-\eta \sum_{s=1}^{t} g_s(i)}$ and $W_t = \sum_{i=1}^{N} W_t(i)$. We have

$$\begin{aligned}
W_t &= \sum_{j=1}^{N} W_{t-1}(j) e^{-\eta g_t(j)} && \leftarrow \quad W_t(j) = W_{t-1}(j) e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^{N} \frac{W_{t-1}(j)}{W_{t-1}} e^{-\eta g_t(j)} && \\
&= W_{t-1} \sum_{j=1}^{N} p_t(j) e^{-\eta g_t(j)} && \leftarrow \quad p_t(j) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(j)}}{\sum_{k=1}^{N} e^{-\eta \sum_{s=1}^{t-1} g_s(k)}} = \frac{W_{t-1}(j)}{W_{t-1}} \\
&\leqslant W_{t-1} \exp\big(-\eta \ell(p_t \cdot x_t, y_t) \big) && \leftarrow \quad \eta\text{-exp-concavity}
\end{aligned}$$

Now, by induction on $t = 1, \dots, T$, this yields using $W_0 = K$

$$W_T \leqslant K \exp\left(-\eta \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) \right) . \tag{4}$$

Proof (Step 2)

On the other hand, upper-bounding the maximum with the sum,

$$\exp\left(-\eta \min_{j\in[K]} \sum_{t=1}^{T} g_t(j)\right) \leqslant \sum_{j=1}^{K} \exp\left(-\eta \sum_{t=1}^{T} g_t(j)\right) \leqslant W_T \,.$$

Combining the above inequality with Inequality (4) and taking the log concludes the proof.

Continuous EWA

Can we obtain a regret with respect to the best combination of experts

$$\min_p \sum_{t=1}^{T} \ell_t(p)$$

instead of the regret with respect to the best fixed expert?

Continuous EWA

$$p_t = \frac{\int_\Theta p e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)} d\mu(p)}{\int_\Theta e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)} d\mu(p)},$$

where μ is the uniform (Lebesgue) measure on $\Theta = \Delta_K$.

Regret bound for continuous EWA

Theorem 4 (Regret of continuous EWA)

Let $T \geqslant 1$. For all sequences of η-exp-concave losses $\ell_1, \dots, \ell_t$ the continuous EWA forecaster satisfies

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \inf_{p \in \Theta} \sum_{t=1}^{T} \ell_t(p) \leqslant \frac{1 + (K-1)\log(T+1)}{\eta}$$

Nice theoretical result but hard to implement because of the integral.

In practice, p_t can be computed by using $(1/T)$-discretization grid of Θ (bad complexity of order T^K!) or by using Monte-Carlo methods to approximate the integral.

Proof (Step 1 – Upper-bound of W_T)

The proof starts similarly to the one of Theorem 2. Let us denote $W_t(p) = e^{-\eta \sum_{s=1}^{t} \ell_s(p)}$, $W_t = \int_\Theta W_t(p) d\mu(p)$ and $d\widehat{\mu}_t(p) = W_t(p) d\mu(p) / W_t$. Then,

$$
\begin{aligned}
W_T &= \int_\Theta e^{-\eta \sum_{t=1}^{T} \ell_t(p)} d\mu(p) \\
&= W_{T-1} \int_\Theta \frac{W_{T-1}(p)}{W_{T-1}} e^{-\eta \ell_T(p)} d\mu(p) \\
&= W_{T-1} \int_\Theta e^{-\eta \ell_T(p)} d\widehat{\mu}_{T-1}(p) && \leftarrow \quad p_T = \int_\Theta p \, d\widehat{\mu}_{T-1}(p) \\
&\leqslant W_{T-1} \exp\big(-\eta \ell_T(p_T)\big) && \leftarrow \quad \eta\text{-exp-concavity} \\
&\leqslant \exp\left(-\eta \sum_{t=1}^{T} \ell_t(p_t)\right), && \leftarrow \quad \text{induction}
\end{aligned}
\tag{5}
$$

Proof (Step 2 – Lower bound of W_T)

For simplicity, we assume that ℓ_t are continuous. Therefore the infimum is a minimum and let $p^* \in \arg\min_{p\in\Theta} \sum_{t=1}^T \ell_t(p)$ and define

$$\Theta_\varepsilon \stackrel{\text{def}}{=} \Big\{(1-\varepsilon)p^* + \varepsilon q, \quad q \in \Theta\Big\}, \qquad \varepsilon \in (0,1)\,.$$

By exp-concavity of ℓ_t, we have for all t and all $p = (1-\varepsilon)p^* + \varepsilon q$

$$e^{-\eta \ell_t(p)} \geqslant (1-\varepsilon)e^{-\eta\ell_t(p^*)} + \varepsilon e^{-\eta\ell_t(q)} \geqslant (1-\varepsilon)e^{-\eta\ell_t(p^*)}$$

Therefore, for all $p \in \Theta_\varepsilon$

$$e^{-\eta\sum_{t=1}^T \ell_t(p)} \geqslant (1-\varepsilon)^T e^{-\eta\sum_{t=1}^T \ell_t(p^*)}$$

Integrating both parts over Θ_ε and using $\mu(\Theta_\varepsilon) = \varepsilon^{K-1}\mu(\Theta)$ (exercise) we get

$$W_T \geqslant \int_{\Theta_\varepsilon} e^{-\eta\sum_{t=1}^T \ell_t(p)} d\mu(p) \geqslant \mu(\Theta)\varepsilon^{K-1}(1-\varepsilon)^T e^{-\eta\sum_{t=1}^T \ell_t(p^*)}\,.$$

Proof (Step 3 – Conclusion)

Combining with (5), using $W_0 = \mu(\Theta)$, taking the log and reorganizing the terms yields

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t) - \sum_{t=1}^{T} \ell_t(p^*) \leqslant \frac{(K-1)\log\frac{1}{\varepsilon} + T\log\frac{1}{1-\varepsilon}}{\eta}\,.$$

Optimizing $\varepsilon = 1/(T+1)$ concludes the proof since

$$T\log\frac{1}{1-\varepsilon} = T\log\left(1+\frac{1}{T}\right) \leqslant 1\,.$$

Next class

Efficient algorithms for general convex losses: online gradient descent, exponentiated gradients.

Prepare a computer with Python installed. You might need to code a little!

References

Thank you!

- Cesa-Bianchi, Nicolo and Gábor Lugosi. Prediction, learning, and games. Cambridge university press, 2006.
- Hazan, Elad et al. "Introduction to online convex optimization". In: Foundations and Trends® in Optimization 2.3-4 (2016), pp. 157–325.
- Lattimore, Tor and Csaba Szepesvári. "Bandit algorithms". In: preprint (2019).
- Shalev-Shwartz, Shai et al. "Online learning and online convex optimization". In: Foundations and Trends® in Machine Learning 4.2 (2012), pp. 107–194.