Lecture 1

More on learning from expert advice

Feb 11, 2005
Lecturer: Anna Karlin
Notes: Ioannis Giotis

1.1 Online experts continued

Consider a sequence of time steps $t = 1 \dots T$, where in each step we choose among m strategies. During each round, our online learner chooses a distribution p^t. Afterwards the environment-adversary chooses a *profit* $\mathcal{P}_i^t$ and *loss* $\mathcal{L}_i^t$ for each strategy i yielding the profit and loss vectors $\mathcal{P}^t$ and $\mathcal{L}^t$. We consider all profit and loss values normalized in the range $[0, 1]$.

The agent earns $(p^t, \mathcal{P}^t - \mathcal{L}^t)$. We also define the *income* on each round as $\mathcal{I}_i^t = \mathcal{P}_i^t - \mathcal{L}_i^t$. $\mathcal{I}_i = sum_{t=1}^T \mathcal{I}_i^t$.

Theorem 1.1. *Given any $\varepsilon \in [0, 1]$, there exists an algorithm for choosing p^t such that*

$$max_i \mathcal{I}_i \leq \mathcal{I} + \varepsilon(\mathcal{P} + \mathcal{L}) + \frac{ln\ m}{\varepsilon}$$

where $\mathcal{I}, \mathcal{P}, \mathcal{L}$ are the expected income, profit and loss.

Let's start by defining $x_i^t = \sum_{\tau=1}^t \mathcal{I}_i^\tau$. Our algorithm is to choose a strategy similarly to the algorithm presented in the last lecture, setting p_i^{t+1} according to $e^{\varepsilon x_i^t}$.

$$p_i^{t+1} = \frac{e^{\varepsilon x_i^t}}{\sum_j e^{\varepsilon x_j^t}}$$

The weights on each strategy are

$$\begin{aligned} w_i^t &= \prod_{\tau=1}^t (1+\varepsilon)^{\mathcal{I}_i^\tau} = \prod_{\tau=1}^t e^{(\mathcal{I}_i^\tau ln(1+\varepsilon))} \\ &= e^{x_i^t \log(1+\varepsilon)} \simeq e^{\varepsilon x_i^t} \end{aligned}$$

The intuition really comes from the continuous (in time) version of the problem. In the continuous version

- $x_i^t =$ cumulative income up to time t.

- $dx_i^t =$ incremental income at time t.
- $$\mathcal{I} = \int_C \sum_{i=1}^{m} p_i^t \; dx_i^t = \int_C \frac{(\sum e^{\varepsilon x_i^t} \; dx_i^t)}{\sum e^{\varepsilon x_j^t}}$$

  , where we're integrating over the curve C of cumulative income of m strategies from $(0, 0, \ldots, 0)$ to $(\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_m)$.

Let

$$\begin{aligned} \Phi(x_1, \ldots, x_m) &= ln \sum_{i=1}^{m} e^{\varepsilon x_i} \\ d\Phi &= \frac{\varepsilon \sum_{i=1}^{m} e^{x_i} \; dx_i}{\sum_{i=1}^{m} e^{\varepsilon x_i}} \end{aligned}$$

We now have

$$\begin{aligned} \mathcal{I} &= \frac{1}{\varepsilon} \int_C d\Phi = \frac{1}{\varepsilon} \left(\Phi(\mathcal{I}_1, \ldots, \mathcal{I}_m) - \Phi(0, \ldots, 0)\right) \\ &= \frac{1}{\varepsilon} \left(ln \sum_{i=1}^{m} e^{\varepsilon \mathcal{I}_i} - ln \; m \right) \\ &\geq \frac{1}{\varepsilon} \left(ln \; max_i e^{\varepsilon \mathcal{I}_i} - ln \; m \right) \\ &= \frac{1}{\varepsilon} max_i ln \; e^{\varepsilon \mathcal{I}_i} - ln \; m \\ &= max_i \mathcal{I}_i - \frac{ln \; m}{\varepsilon} \end{aligned}$$

We conclude that

$$max_i \mathcal{I}_i \leq \mathcal{I} + \frac{ln \; m}{\varepsilon}$$

The proof for the discrete version of the problem is similar.

Some concluding remarks about this algorithm. We're interested in finding $max_i \alpha_i$. $ln \; \sum_i^m e^{\alpha_i}$ is a very good approximation since the following holds

$$max_i \alpha_i \leq ln \; \sum_i^m e^{\alpha_i} \leq max_i \alpha_i + ln \; m$$

Furthermore, this function is differentiable

$$\frac{\mathfrak{d} \; ln \; \sum e^{\alpha_i}}{\mathfrak{d} \; \alpha_i} = \frac{e^{alpha_i}}{\sum e^{\alpha_j}}$$

Intuitively, it is some form of steepest decent, going in the direction of the partial derivatives.

1.2 Minimizing regret

We have a decision maker taking N actions, choosing a probability distribution p^t on each round. The loss is defined as $l^t \in [0,1]^N$ and the total loss is $L_H = \sum_t \sum_i p_i^t l_i^t$.

So far we were minimizing *external* regret

$$L_H - min_i L_i^t$$

Minimizing *internal* regret is

$$L_H - L_{min}(i \to j) = max_{i,j} \sum_t p_i^t (l_i^t - l_j^t)$$

where $(i \to j)$ represents making one global change in the strategy.

Minimizing *swap* regret is

$$L_H^T - min\ L^T(i \to F(i)) = \sum_t \sum_i p_i^t (l_i^t - l_{F(i)}^t)$$

where F is a function $\{1 \dots N\} \to \{1 \dots N\}$.

1.2.1 Correlated equilibria

Definition 1.1. The empirical distribution over A_j is

$$p(\alpha_1, \dots, \alpha_n) = \frac{1}{T} \sum_{t=1}^{T} p_1^t(\alpha_1) p_2^t(\alpha_2) \cdots p_n^t(\alpha_n)$$

Definition 1.2. A probability distribution p over a set of actions $A_1 \times A_2 \times \cdots \times A_n$ is an ε-correlated equilibrium if $\forall j, \forall F : A_j \to A_j$

$$E_{\alpha \sim p}(u_j(\alpha_j, \alpha_{-j})) \leq E_{\alpha \sim p}(u_j(F(\alpha_j), \alpha_{-j})) + \varepsilon \tag{1.1}$$

$u_j(\alpha_i, \dots, \alpha_m)$ is the loss to player j when $\forall k$, player k plays α_k.

(1.1) can be written alternatively as

$$\sum_{(\alpha_1, \dots, \alpha_n)} p(\alpha_1, \dots, \alpha_n) u_j(\alpha_1, \dots, \alpha_n) \leq \sum_{(\alpha_1, \dots, \alpha_n)} p(\alpha_1, \dots, \alpha_n) u_j(\dots, F(\alpha_j), \dots) + \varepsilon$$

Theorem 1.2. *Consider a game of n players, where for T times steps, each player plays according to some strategy with swap regret $\leq \alpha$. Then the empirical distribution of joint actions is a $\frac{\alpha}{T}$-correlated equilibrium.*

Proof. Canceling out T, we need to show that

$$\sum_t \sum_{(\alpha_1, \dots, \alpha_n)} p_1^t(\alpha_1) p_2^t(\alpha_2) \cdots p_n^t(\alpha_n) u_j(\alpha_1, \dots, \alpha_n) \leq$$
$$\sum_t \sum_{(\alpha_1, \dots, \alpha_n)} p_1^t(\alpha_1) p_2^t(\alpha_2) \cdots p_n^t(\alpha_n) u_j(\dots, F(\alpha_n), \dots) + \alpha$$

The left hand side can be written as

$$\sum_t \sum_{\alpha_j} p_j^t(\alpha_j) \sum_{\alpha_{-j}} p_{-j}^t(\alpha_{-j}) u_j(\alpha_1, \ldots, \alpha_n)$$

$\square$