15-859E: Advanced Algorithms **CMU, Spring 2015**
Lecture #13: Applications of MW: Zero-Sum Games February 11, 2015
Lecturer: Anupam Gupta Scribe: Eugene Choi

1 Last Time

Last time, we covered a multiplicative weights algorithm for learning with small regret.

Let $\ell^t \in [-1,1]^N$ be a "loss vector." Define $\Delta_N = \{x \in [0,1]^N \mid \sum_{i=1}^N x_i = 1\}$ to be the N-dimensional probability simplex.

Last time, we showed:

Theorem 13.1. *For every $0 < \epsilon \le 1$, there exists an algorithm Hedge(ϵ) such that for all times $T > 0$, for every sequence of loss vectors $(\ell^1, \ldots, \ell^T)$, and for every $i \in \{1, \ldots, n\}$, at every time $t \le T$, Hedge(ϵ) produces $p^t \in \Delta_N$ such that*

$$\sum_{t=1}^{T} \langle \ell^t, p^t \rangle \le \sum_{t=1}^{T} \langle \ell^t, e_i \rangle + \epsilon T + \frac{\ln N}{\epsilon},$$

where e_i is the ith vector in the standard basis of $\mathbb{R}^N$. Note that the first term on the right hand side represents the loss of the ith expert, and the last two terms represents the regret of not having always chosen the ith expert.

Note that if we choose $\epsilon = \sqrt{\frac{\ln N}{T}}$, then $\epsilon T + \frac{\ln N}{\epsilon} = 2\sqrt{T \ln N}$, so that the regret term is *sublinear* in time T. This indicates that the average regret of Hedge(ϵ) converges towards the best expert, so that Hedge(ϵ) is in some sense "learning".

For future reference, we state the analogous result for gains g^t instead of losses ℓ^t, i.e., $g^t = -\ell^t$.

Theorem 13.2. *For every $0 < \epsilon \le 1$, there exists an algorithm $Hedge_g(\epsilon)$ such that for all times $T > 0$, for every sequence of gain vectors $(g^1, \ldots, g^T)$, and for every $i \in \{1, \ldots, n\}$, at every time $t \le T$, $Hedge_g(\epsilon)$ produces $p^t \in \Delta_N$ such that*

$$\sum_{t=1}^{T} \langle g^t, p^t \rangle \ge \sum_{t=1}^{T} \langle g^t, e_i \rangle - \epsilon T - \frac{\ln N}{\epsilon},$$

where e_i is the ith vector in the standard basis of $\mathbb{R}^N$. Note that the first term on the right hand side represents the gain of the ith expert, and the last two terms represents the regret of not having always chosen the ith expert.

We also state a corollary of 13.2 that we will use.

Corollary 13.3. *Let $\rho \ge 1$. For every $0 < \epsilon \le \frac{1}{2}$, for all times $T \ge \frac{4\rho^2 \ln N}{\epsilon^2}$, for all sequences of gain vectors $(g^1, \ldots, g^T)$ with each $g^t \in [-\rho, \rho]^N$, and for all $i \in \{1, \ldots, N\}$, at every time $t \le T$, $Hedge_g(\epsilon)$ produces $p^t \in \Delta_N$ such that*

$$\frac{1}{T} \sum_{t=1}^{T} \langle g^t, p^t \rangle \ge \frac{1}{T} \sum_{t=1}^{T} \langle g^t, e_i \rangle - \epsilon.$$

2 Zero-Sum Games

A zero sum game can be described by a pay-off matrix $M \in \mathbb{R}^{m \times n}$, where the two players are the "row player" and the "column player". Simultaneously, the row player chooses a row i and the column player chooses a column j, and the row player receives a pay-off of $M_{i,j}$. Alternatively, the column player loses $M_{i,j}$; hence, the name "zero-sum".

Given a strategy $x \in \Delta_m$ for the row player and a strategy $y \in \Delta_n$ for the column player, the expected pay-off to the row player is

$$\mathbf{E}[\text{pay-off to row}] = x^T M y.$$

The row player wants to maximize this value, while the column player wants to minimize this value. Note that the choices of x and y are made simultaneously. Later, we will see that these choices really do not have to be made at the same time.

Suppose row player fixes a strategy $x \in \Delta_m$. Knowing the row player's strategy, the column player can choose a response to minimize the row player's expected winnings:

$$C(x) = \min_{y \in \Delta_n} x^T M y = \min_{j \in [n]} x^t M e_j.$$

The equality holds because if the column player already knows the row player's strategy (x), then the column player's best strategy is to choose the column that minimizes the row player's expected winnings.

Analogously, suppose the column player fixes a strategy $y \in \Delta_n$. Knowing the column player's strategy, the row player can choose a response to maximize his own expected winnings:

$$R(y) = \max_{x \in \Delta_m} x^T M y = \max_{i \in [m]} e_i^T M y.$$

Overall, the row player wants to achieve $\max_{x \in \Delta_m} C(x)$, and the column player wants to achieve $\min_{y \in \Delta_n} R(y)$.

Theorem 13.4. *(Von Neumann's Minimax) For any finite zero-sum game $M \in [-1,1]^{m \times n}$,*

$$\max_{x \in \Delta_m} C(x) = \min_{y \in \Delta_n} R(y).$$

The common value V is called the value of the game M.

Proof. We treat each row of M as an expert. At each time step t, the row player produces $p^t \in \Delta_m$. Initially, $p^1 = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, which represents that the row will choose any row with equal probability when he has no information to work with.

At each time t, the column player plays the best response to p^t, i.e., $j_t := \arg\max_{j \in [n]} (p^1)^T M e_j$. Let the gain vector at time t be $g^t := M e_{j_t} \in [-1,1]^m$. Then the pay-off to the row player at time t is $\langle g^t, p^t \rangle = (p^t)^T M e_{j_t} = C(p^t)$.

Define $\hat{x} := \frac{1}{T} \sum_{t=1}^{T} p^t$ and $\hat{y} := \frac{1}{T} \sum_{t=1}^{T} e_{j_t}$.

Claim 13.5. $C(\hat{x}) \le R(\hat{y}) \le C(\hat{x}) + \epsilon$ *if* $T \ge \frac{4 \ln N}{\epsilon^2}$.

Proof. We first show that for any $x \in \Delta_m$ and $y \in \Delta_n$, $C(x) \le R(y)$. This is because the row can only do at least as well if he goes second (after the column player) than if he goes first (before the column player). Formally,

$$C(x) = \min_{y' \in \Delta_n} x^T M y' \le x^T M y \le \max_{x' \in \Delta_m} (x')^T M y = R(y).$$

We then show that $R(\hat{y}) \le C(\hat{x}) + \epsilon$. The pay-off to the row player at time t is $\langle g^t, p^t \rangle = C(p^t)$. So the average pay-off is

$$\frac{1}{T}\sum_{t=1}^{T} \langle p^t, g^t \rangle = \frac{1}{T}\sum_{t=1}^{T} C(p^t) \le C\left(\frac{1}{T}\sum_{i=1}^{T} p^t\right) = C(\hat{x}),$$

where the inequality follow's from the convexity of C (i.e., min is a convex function).

By (13.3), we also have for every $i \in [m]$,

$$\frac{1}{T}\sum_{t=1}^{T} \langle p^t, g^t \rangle \ge \frac{1}{T}\sum_{t=1}^{T} \langle e_i, g^t \rangle - \epsilon = \left\langle e_i, \frac{1}{T}\sum_{t=1}^{T} g^t \right\rangle - \epsilon = \langle e_i, M\hat{y} \rangle - \epsilon.$$

Hence, it follows that

$$\frac{1}{T}\sum_{t=1}^{T} \langle p^t, g^t \rangle \ge \max_{i \in [m]|} \langle e_i, M\hat{y} \rangle - \epsilon = R(\hat{y}) - \epsilon.$$

□

With the claim, for every $n \ge 2$, we can find $\hat{x}_n$ and $\hat{y}_n$ such that $|C(\hat{x}) - R(\hat{y})| \le \frac{1}{n}$. If we define $F(x, y) = C(x) - R(y)$, then $|F(\hat{x}_n, \hat{y}_n)| \le \frac{1}{n}$. Since $\Delta_m \times \Delta_n$ is bounded, (by Bolzano-Weierstrass) we can find a convergent subsequence of $(\hat{x}_n, \hat{y}_n)$. Since $\Delta_m \times \Delta_n$ is compact, that subsequence converges to some $(\hat{x}, \hat{y}) \in \Delta_m \times \Delta_n$, so that $F(\hat{x}, \hat{y}) = 0$. This shows that we can find $\hat{x}$ and $\hat{y}$ such that $C(\hat{x}) = R(\hat{y})$, which proves the theorem. □