CS 6840 Algorithmic Game Theory September 13, 2024

Lecture 8: Learning in 2-person zero-sum games

Instructor: Eva Tardos *Scribe: Richard Niu*

Today's lecture covers what happens when learning algorithms are applied to two-person zero-sum games. Specifically, in this context, learning players can find a Nash equilibrium; in fact, this is closely associated and related to the idea of duality in linear programming.

1 Introduction

Recall that two-person zero-sum games can be described by a single matrix A, which the payoffs indicated are for the row player and are also thus the loss for the column player. For example, the following matrix describes a game of rock-paper-scissors:

	R	P	S
R	0	-1	$+1$
P	$+1$	0	-1
S	-1	$+1$	0

Let $\mathbf{x}$ and $\mathbf{y}$ denote probability vectors for the row and column players, respectively. Then the expected payoff of player 1 (the row player) is equal to

$$\sum_{i,j} \mathbf{x}_i \mathbf{y}_j A_{ij} = \mathbf{x}^\intercal A \mathbf{y}.$$

Definition 1. A *Nash equilibrium* occurs between two vectors $\mathbf{x}$ and $\mathbf{y}$ if

$$\begin{aligned} \mathbf{x}^\intercal A \mathbf{y} &\geq \bar{\mathbf{x}}^\intercal A \mathbf{y} && \text{for all probability vectors } \bar{\mathbf{x}}, \text{ and} \\ \mathbf{x}^\intercal A \mathbf{y} &\leq \mathbf{x}^\intercal A \bar{\mathbf{y}} && \text{for all probability vectors } \bar{\mathbf{y}}. \end{aligned}$$

Let's consider the implications of what happens when one player goes first, and the other goes second. In this context, "going" means committing to a mixed strategy/probability distribution.

If the row player, 1 (represented by x) goes first, they wish to find $\max_{\mathbf{x}} \min_{\mathbf{y}} \mathbf{x}^\intercal A \mathbf{y}$; that is, they should play the strategy that maximizes their payoff, knowing that y will play their best-response function. Similarly, if player 1 goes second, then their payoff will be $\min_{\mathbf{y}} \max_{\mathbf{x}} \mathbf{x}^\intercal A \mathbf{y}$; player 2 will want to minimize their loss, given player 1 plays their best-response function.

It is clear that going second should be advantageous for player 1, since their strategy space when going first remains intact when going second, and they only receive additional information after player 2 acts when compared to before; that is, $\max_{\mathbf{x}} \min_{\mathbf{y}} \mathbf{x}^\intercal A \mathbf{y} \leq \min_{\mathbf{y}} \max_{\mathbf{x}} \mathbf{x}^\intercal A \mathbf{y}$.

2 Learning

We'll now investigate what happens when no-regret learning strategies are employed in this game. (A slight caveat: we'll take "no-regret" to mean literally no regret at all, while in actuality "real" no-regret learning has some tiny regret as $t \to \infty$.)

Suppose this game is iterated over time periods $1 \dots t$, with player x playing the vectors $\mathbf{x}^1 \dots \mathbf{x}^t$ and player y playing the vectors $\mathbf{y}^1 \dots \mathbf{y}^t$. We can quantify the *expected* loss of the players as follows:

- Column player's (expected) loss at time t for a pure strategy j: $\ell_t^c(j) = \sum_i \mathbf{x}_i^t A_{ij}$
- Row player's (expected) loss at time t for a pure strategy i: $\ell_t^r(i) = -\sum_j A_{ij}\mathbf{y}_j^t$

We now consider the column player's average loss over T periods, which is:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}^t)^\intercal A\mathbf{y}^t &\leq \frac{1}{T}\min_j \sum_{t=1}^{T}\ell_t^c(j) \qquad \text{(no-regret assumption/"best strategy in hindsight")}\\ &= \frac{1}{T}\min_j \sum_t \sum_i \mathbf{x}_i^t A_{ij}\\ &= \min_j \sum_i \bar{\mathbf{x}}_i A_{ij}\end{aligned}$$

where we define $\bar{\mathbf{x}} = \frac{1}{T}\sum_t x^t$.

Similarly, the average loss of the row player satisfies

$$\begin{aligned}-\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}^t)^\intercal A\mathbf{y}^t &\leq \frac{1}{T}\min_i \sum_{t=1}^{T}\ell_t^r(i)\\ &= \frac{1}{T}\min_i -\sum_t \sum_j A_{ij}\mathbf{y}_j^t\\ &= \min_i -\sum_j A_{ij}\bar{\mathbf{y}}_j\end{aligned}$$

where $\bar{\mathbf{y}}$ is defined similarly.

Claim: Choices of $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ constitute a Nash equilibrium.

Proof: Next time.