CS861: Theoretical Foundations of Machine Learning Lecture 27 - 11/06/2024
University of Wisconsin–Madison, Fall 2024

Lecture 27: Learning in games (cont'd)

Lecturer: Kirthevasan Kandasamy *Scribed by: Hongyi Huang, Shuqi Bi*

Continuing from the end of last time, we will keep talking about game theory and its connection to online learning. We will start by introducing **two-player normal form game**. Then, we will talk about two solution concepts: **Nash equilibrium** and **Safety strategies**. And we will see how we can use online learning method to establish the proof of **Minimax theorem**.

Definition 1 (Normal Form Game)**.** *In a finite two-player NFG, player 1 has m available actions and player 2 has n available actions. If player 1 chooses $a_1 \in [m]$ and player 2 chooses $a_2 \in [n]$, they receive payoffs $Q^{(1)}(a_1, a_2), Q^{(2)}(a_1, a_2)$ respectively. Here, $Q^{(1)}, Q^{(2)} \in \mathbb{R}^{m\times n}$.*

Example 1 (Rock-Paper-Scissors)**.** Let's consider the Rock-Paper-Scissors game. The following is the payoff (or utility) table for Player 1 and Player 2.

(P1,P2)	R	P	S
R	(0,0)	(-1,1)	(1,-1)
P	(1,-1)	(0,0)	(-1,1)
S	(-1,1)	(1,-1)	(0,0)

(Question) How to Choose a Strategy/Action for each player?
Thoughts: A player's best action depends on the other player's action.

Example 2 (Deer/Stag Hunt)**.** There are two players, who both go out hunting. Here, two players must independently choose whether to hunt a deer or a rabbit. Hunting a deer requires both players' cooperation to succeed and offers a greater reward, but also carries risk since each hunter depends on the other to commit. Hunting a rabbit, on the other hand, can be done individually and is less risky but provides a smaller reward. The following is the reward table:

	D	R
D	(4,4)	(0,2)
R	(2,0)	(1,1)

Example 3 (Driver vs. Inspector)**.** Consider driver and inspector as two opposing players. Driver can choose to drive "legally" or "illegally". And inspector can choose "don't inspect" or "inspect". The following is the payoff table:

	Don't Inspect	Inspect
Legal	(0,0)	(0,-1)
Illegal	(10,-10)	(-90,-6)

Definition 2 (Mixed Strategies)**.** *Instead of choosing an action (pure strategy), players can choose distributions over the actions. For example, $P_1 : x \in \Delta([m]), P_2 : y \in \Delta([n])$. Later, we can think of all strategy as mixed strategy.*

With above definition, we can consider the expected payoffs of x, y.

$$P_1 : x^T Q^{(1)} y \qquad P_2 : x^T Q^{(2)} y$$

Several solution concepts arise from the strategies of two players. We will look specifically at 2 of them:

- Nash equilibrium
- Safety strategy

Definition 3 (Nash equilibrium)**.** *A strategy x for player 1 is a best response to strategy y of player 2, if it maximizes player 1's utility. i,e.,*

$$x^T Q^{(1)} y \geq x'^T Q^{(2)} y \quad , \quad \forall x' \in \Delta([m])$$

Similarly, for player 2, y is a best response to x if

$$x^T Q^{(2)} y \geq x^T Q^{(1)} y' \quad , \quad \forall y' \in \Delta([n])$$

*A pair (x_*, y_*) is a **Nash Equilibrium** (NE) if: x_* is player 1's best response to y_* and y_* is player 2's best response to x_*. That is, no player can unilaterally deviate and do better.*

Here are some NE choices for previous examples:

1. Rock-Paper-Scissors:
   $x_* = y_* = (1/3, 1/3, 1/3)$ is a NE.
2. Deer/Stag Hunt:
   $x_* = y_* = (1, 0)$, (D,D)
   $x_* = y_* = (0, 1)$, (R,R)
   $x_* = y_* = (1/3, 2/3)$ is a mixed NE.
3. Driver vs. Inspector:
   $x_* = (0.8, 0.2), \ y_* = (0.9, 0.1)$

Definition 4 (Safety Strategies)**.** *Define $g_1(x) \triangleq \min_{y \in \Delta([n])} x^T Q^{(1)} y$, $g_2(y) = \min_{x \in \Delta([m])} x^T Q^{(2)} y$. Then P1 and P2's safety strategies are:*

$$x^* = \arg \max_{x \in \Delta([m])} g_1(x) = \arg \max_x \min_y \ x^T Q^{(1)} y$$

$$y^* = \arg \max_{y \in \Delta([n])} g_2(y) = \arg \max_y \min_x \ x^T Q^{(2)} y$$

Some safety strategy choice for previous examples:

- RPS:

$$x_* = y_* (1/3, 1/3, 1/3)$$

- Driver vs. Inspector:

$$x_* = (1, 0), \quad y_* = (0, 1)$$

(Question): When are Safety Strategies always equal to NE ?
Answer: Two-player zero-sum games (TPZSG)

Definition 5 (Two-player zero sum game (TPZSG))**.** *In a two-player zero-sum-game,*

$$Q^{(1)} = -Q^{(2)} \triangleq Q$$

- NE in TPZSG:

$$x^T Q y \leq x_*^T Q y_* \leq x_*^T Q y \qquad \forall x \in \Delta([m]), y \in \Delta([n])$$

- Safety strategy in TPZSG:

$$(1) : \tilde{x} = \arg\max_x \min_y x^T Q y$$

$$(2) : \tilde{y} = \arg\min_y \max_x x^T Q y$$

Interpretation of $\tilde{x}, \tilde{y}$:
P1 has to announce their strategy first. P1 knows P2 will choose $\arg\min_y x^T Q y$ if she announces x. Therefore, she will choose:

$$\tilde{x} = \arg\max_x \min_y \quad x^T Q y$$

P1's expected payoff (utility):

$$(*) : \max_x \min_y x^T Q y$$

Similarly, if P2 announces first, P2's expected utility is:

$$-\min_y \max_x x^T Q y$$

P1's expected utility is:

$$(**) : \min_y \max_x x^T Q y$$

The following is a brief argument to show that $(*) \leq (**)$:

For any continuous function $f(x, y)$,

$$g_1(x) \triangleq \min_y f(x, y)$$

$$g_2(y) \triangleq \max_x f(x, y)$$

If g_1 and g_2 are also continuous, we have:

$$\min_y f(x, y) \leq f(x, y') \quad \forall y' \in \Delta([n])$$

$$\max_x \min_y f(x, y) \leq \max_x f(x, y') \quad \forall y'$$

$$\max_x \min_y f(x, y) \leq \min_{y'} \max_x f(x, y')$$

We can also show $(*) = (**)$. We will use the following theorem:

Theorem 4 (Von-Neumann's MiniMax)**.** *In any TPZSG,*

$$(*) = \max_x \min_y x^T Q y = \min_y \max_x x^T Q y = (**)$$

This value, denoted $V(Q)$, is called the value of the game.

Proof: Since we have shown $(*) \leq (**)$ before, we will show $(*) \geq (**)$ in the following.
Consider the following multi-round game:

- On each round, P_2 chooses a mixed strategy $y_t \in \Delta([n])$, via some policy π.
- After this, P_1 chooses the best response

$$x_t = \arg\max_{x \in \Delta([n])} x^T Q y_t$$

- We will define P_2's regret as

$$R_T(\pi, E) = \sum_{t=1}^{T} x_t^T Q y_t - \min_{y \in \Delta([n])} \sum_{t=1}^{T} x_t^T Q y$$

  where E refers to the estimate: Q, P_1's behavior, what is known to P_2.

Claim 1. *If there exists a policy π which achieves sublinear regret, then*

$$\max_x \min_y x^T Q y \geq \min_y \max_x x^T Q y$$

Proof of the claim:
Let $x_1, ..., x_T, y_1, ..., y_T$ be the strategies used by both players on each round. Let

$$\bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t$$

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

$$\begin{aligned}
\min_y \max_x x^T Q y &\leq \max_x x^T Q \bar{y} \\
&= \max_x \frac{1}{T} \sum_{t=1}^{T} x^T Q y_t \\
&= \frac{1}{T} \sum_{t=1}^{T} \max_x x^T Q y_t \\
\text{(by choice/definition of } x_t) &= \frac{1}{T} \sum_{t=1}^{T} x_t^T Q y_t \\
&= \frac{1}{T} \left(\min_{y \in \Delta([n])} \sum_{t=1}^{T} x_t^T Q y + R_T(\pi, E) \right) \\
&= \min_{y \in \Delta([n])} \bar{x}^T Q y + \frac{1}{T} R_T(\pi, E) \\
&\leq \underbrace{\max_x \min_y x^T Q y}_{(**)} + \underbrace{\frac{1}{T} R_T(\pi, E)}_{\to 0 \text{ as } T \to \infty} \quad \text{(by claim assumption)}
\end{aligned}$$

So, $\forall \epsilon > 0$, we can find any policy π s.t. $\frac{R}{T} \leq \epsilon$. Then, $(*) \leq (**)$