Introduction to Artificial Intelligence

Lecture 18: Games and strategies

November 6, 2025

Lecture plan

- **Games and strategies**
  - **Zero-sum game**
  - None-zero-sum game: Nash equilibrium
  - Cooperative games

Recap: Two-player zero-sum games

- Players = {agent, opponent}

- **Definition: two-player zero-sum game**
  - s_{start}: starting state
  - Action(s): possible actions from state s
  - Successor(s, a): resulting state if choose action a in state s
  - IsEndState(s): whether s is an end state
  - Utility(s): agent's utility for end state s
  - Player(s) $\in$ Players: player who controls state s

Two-finger Morra

- Players A and B each show 1 or 2 fingers
  - If both show 1, B gives A 2 dollars
  - If both show 2, B gives A 4 dollars
  - Otherwise, A gives B 3 dollars

Two-finger Morra

- What are the possible actions?
  - Player A chose 1, player B chose 1
  - Player A chose 1, player B chose 2
  - Player A chose 2, player B chose 1
  - Player A chose 2, player B chose 2

Payoff matrix

- Definition: single-move simultaneous game
  - Players = $\{A, B\}$
  - Actions: all possible actions
  - $V(a, b)$: A's utility if A chooses action a, B choose b (Let V be the payoff matrix)

- Example: two-finger Morra payoff matrix

Player A	1 finger	2 fingers
1 finger	2	-3
2 fingers	-3	4

Player B	1 finger	2 fingers
1 finger	-2	3
2 fingers	3	-4

Strategies

- Definition:
  - **Pure strategy**: a pure strategy is a single action: $a \in Actions$
  - **Mixed strategy**: a mixed strategy is a probability distribution over all possible actions

$$0 \leq \pi(a) \leq 1 \quad \text{for } a \in Actions$$

- Example: two-finger Morra strategies
  - Always play action 1: $\pi = [1, 0]$
  - Always play action 2: $\pi = [0, 1]$
  - Uniformly chosen at random: $\pi = \left[\frac{1}{2}, \frac{1}{2}\right]$

Game evaluation

- Definition: game evaluation
  - The value of the game if player A follows π_A and player B follows π_B is

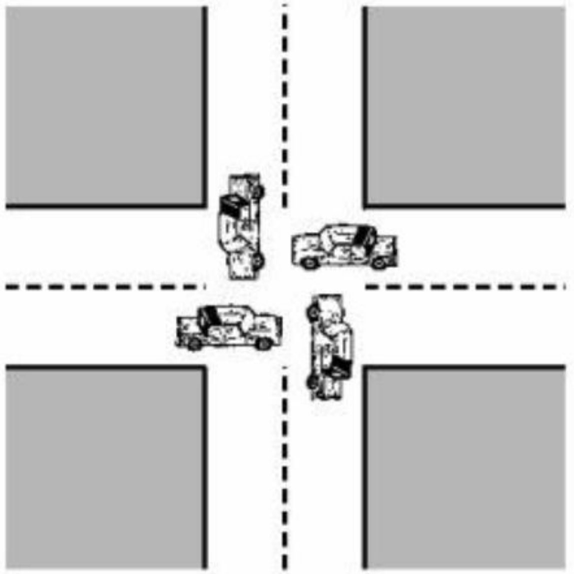
$$V(\pi_A, \pi_B) = \sum_{a,b} \pi_A(a) \pi_B(b) V(a, b)$$

How to optimize the game value?

- Game value:

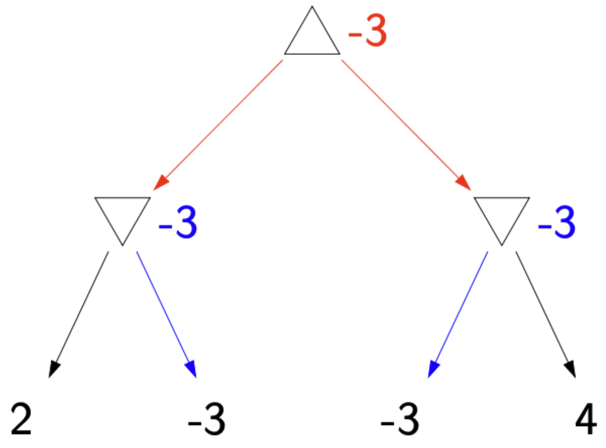
$$V(\pi_A, \pi_B)$$

- Challenge: player A wants to **maximize**, whereas player B wants to minimize



Pure strategies: Who goes first?

- Player A goes first:



Going second is no worse!

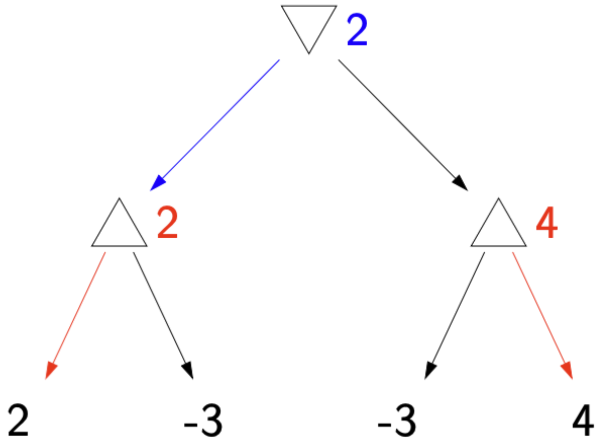
$$\max_a \min_b V(a,b) \le \min_b \max_a V(a,b)$$

Proof: For any fixed a, b

$$\min_{b'} V(a,b') \le V(a,b) \le \max_{a'} V(a',b)$$

Now we may take max on the left inequality, and min on the right inequality, to conclude the minimax inequality

- Player B goes first:



Mixed strategies

- Player A reveals: $\pi_A = \left[\frac{1}{2}, \frac{1}{2}\right]$
- Value $V(\pi_A, \pi_B) = \pi_B(1) \times \left(-\frac{1}{2}\right) + \pi_B(2) \times \left(\frac{1}{2}\right)$
- Optimal strategy for player B is $\pi_B = [1, 0]$, it's a pure strategy!

- **Claim:** For any fixed mixed strategy π_A:

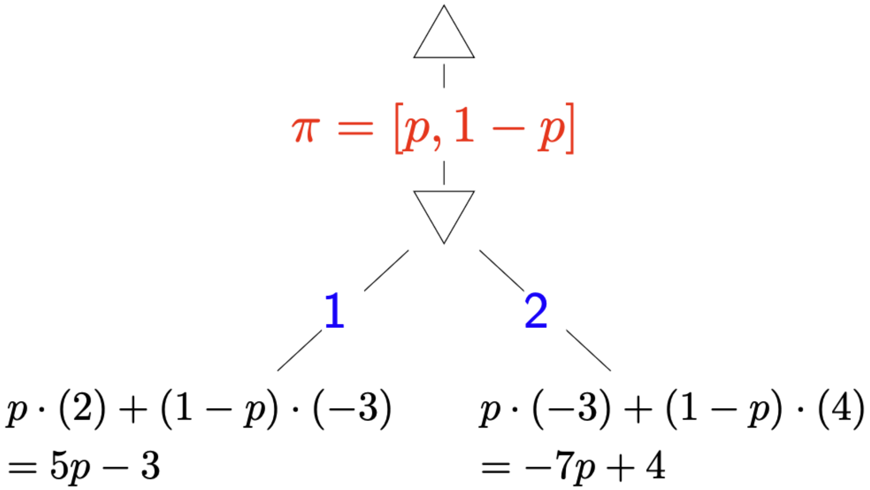
$$\min_{\pi_B} V(\pi_A, \pi_B)$$

can be attained by a pure strategy π_B

Can you convince yourself that this statement is correct?

Mixed strategies

- Player *A* first reveals his or her mixed strategy

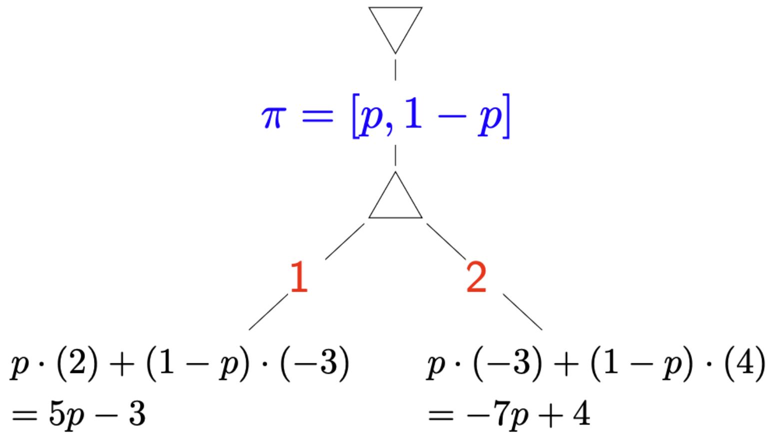


- Player B chooses the minimum of the two, leading to the maximin value of game:

$$\max_{0 \le p \le 1} \min\{5p - 3, -7p + 4\} = -\frac{1}{12} \text{ (with } p = \frac{7}{12}\text{)}$$

Mixed strategies

- Player B first reveals his or her mixed strategy



- Player A chooses the maximum of the two, leading to the minimax value of game:

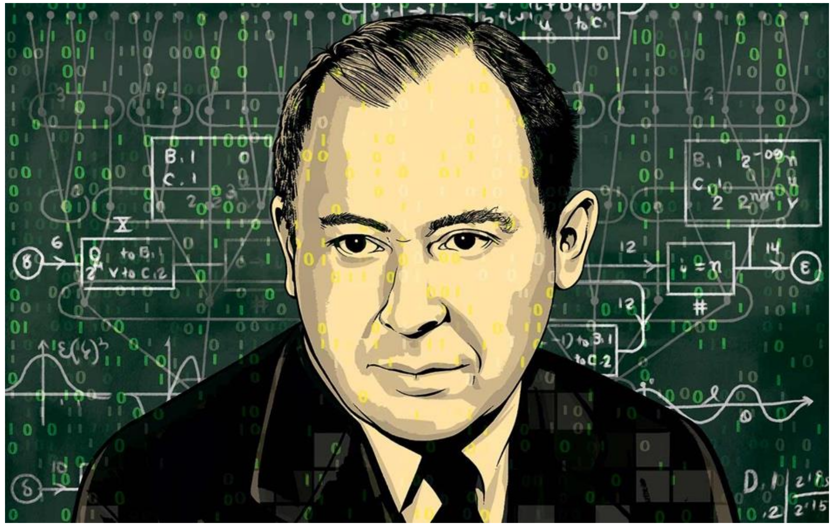
$$\min_{p \in [0,1]} \max\{5p - 3, -7p + 4\} = -\frac{1}{12} \text{ (again with } p = \frac{7}{12}\text{)}$$

Minimax theorem

- Theorem: the minimax theorem [John von Neumann, 1928]
  - For every simultaneous two-player zero-sum game with a finite number of actions

$$\max_{\pi_A \in \Delta_m} \min_{\pi_B \in \Delta_n} V(\pi_A, \pi_B) = \min_{\pi_B \in \Delta_n} \max_{\pi_A \in \Delta_m} V(\pi_A, \pi_B),$$

where π_A, π_B range over mixed strategies over a finite set



Hungarian-American mathematician and physicist, 1903 - 1957

Proof of minimax theorem

- Let M be the payoff matrix for player A. Suppose A and B use mixed strategies $x \in \Delta_m$ and $y \in \Delta_n$, where Δ_m and Δ_n refer to the simplex on the unit sphere in the positive orthant

- The expected payoff for player A is
$$f(x, y) = x^\top M y$$

- First, fix x. Because $x^\top M$ is linear, its minimum is attained at a vertex, i.e., at some pure column e_j. Thus
$$\min_{y \in \Delta_n} f(x, y) = \min_{1 \le j \le n} x^\top M e_j$$

Proof of minimax theorem

- Define $F(x) = \min_{1 \le j \le n} x^\top M e_j$. F is the pointwise minimum of linear functions. One can verify that
  - $F(x)$ is concave
  - $F(x)$ is continuous in the simplex Δ_m
- Therefore, there exists some $x^* \in \Delta_m$ that maximizes $F(x)$. Let

$$v := F(x^*) = \min_{1 \le j \le n} x^{*\top} M e_j$$

- Consider the set of "tight" columns $J = \{j \colon x^{*\top} M e_j = v\}$. Take any probability vector y^* supported on J. Then

$$x^{*\top} M y^* = v$$

because every column in the support gives payoff v to x^*

Proof of minimax theorem

- For every $y \in \Delta_m$, $F(x^*) \le f(x^*, y)$. With $y = y^*$, we get $f(x^*, y^*) = v$. Hence
$$\max_{x\in\Delta_m} \min_{y\in\Delta_n} f(x,y) \ge \min_{y\in\Delta_n} f(x^*, y) = v = f(x^*, y^*)$$

- Next, we have that
$$\min_{y\in\Delta_n} \max_{x\in\Delta_m} f(x,y) \le \max_{x\in\Delta_m} f(x, y^*) \le v = \max_{x\in\Delta_m} F(x)$$

The second part is because on every $j \in J$, $f\left(x^\star, e_j\right) = v$

- Since $\max_{x\in\Delta_m} \min_{y\in\Delta_n} f(x,y) \ge v$, we have
$$\max_{x\in\Delta_m} \min_{y\in\Delta_n} x^\top M y \ge \min_{y\in\Delta_n} \max_{x\in\Delta_m} x^\top M y$$

Proof of minimax theorem

- Finally, recall the minimax inequality, which asserts that

$$\max_{\pi_A} \min_{\pi_B} V(\pi_A, \pi_B) \leq \min_{\pi_B} \max_{\pi_A} V(\pi_A, \pi_B)$$

Combined together, we thus complete the proof

Lecture plan

- **Games and strategies**
  - Zero-sum game
  - **None-zero-sum game: Nash equilibrium**
  - Cooperative games

Prisoner's dilemma

- Prosecutor asks A and B individually if each will testify against the other
  - If both testify, then both are sentenced to 5 years in jail
  - If both refuse, then both are sentenced to 1 year in jail
  - If only on testifies, then he or she gets out for free; the other gets a 10 year sentence

Prisoner's dilemma

- What are the possible outcome?
  - Player *A* testified, player *B* testified
  - Player *A* testified, player *B* refused
  - Player *A* refused, player *B* testified
  - Player *A* refused, player *B* refused

Prisoner's dilemma

- Payoff matrix

	testify	refuse
testify	$A = -5, B = -5$	$A = -10, B = 0$
refuse	$A = 0, B = -10$	$A = -1, B = -1$

- Let $V_p(\pi_A, \pi_B)$ be the utility for player p

Prisoner's dilemma

- If both players had refused, then one of the players could testify to improve his/her payoff (from -1 to 0)
- If one player testified, then the other had to choose testified, otherwise he/she would be in the jail for 10 years instead of 5 years
- This is not the highest possible reward, but it is stable in the sense that neither player would want to change his or her strategy

- This kind of situation is commonly known as the **Nash equilibrium**

Nash equilibrium

- Different from zero-sum games, a Nash equilibrium is a kind of fixed-point state, where no player has any incentive to change his or her policy unilaterally

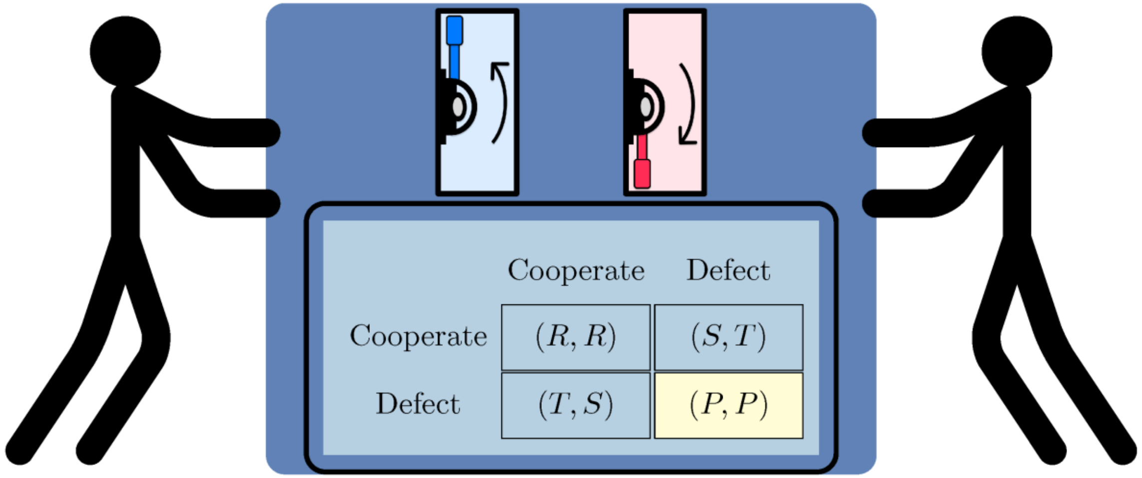
- **Definition:** A Nash equilibrium is any kind of policies (π_A^*, π_B^*) such that no player has an incentive to change his or her strategy, conditioned on the other player's strategy

$$V_A(\pi_A^*, \pi_B^*) \geq V_A(\pi_A, \pi_B^*) \text{ for all } \pi_A$$

$$V_B(\pi_A^*, \pi_B^*) \geq V_B(\pi_A^*, \pi_B) \text{ for all } \pi_B$$

Nash equilibrium

- Since the game is no longer a zero-sum game, we cannot apply the minimax theorem, but we can still get a weaker result
- **Theorem:** Nash's existence theorem [1950]
  - In any finite-player game with finite number of actions, there exits **at least one (mixed-strategy)** Nash equilibrium



  - Proof uses Kakutani's fixed-point theorem (skipped)

Examples of Nash equilibria

- The minimax strategies for zero-sum are also equilibria and they are simultaneously the global optima (exercise question)
- **Example 1:** Two-finger Morra
  - Nash equilibrium: A and B both play $\pi = \left[\frac{7}{12}, \frac{5}{12}\right]$

$$\min_{p \in [0,1]} \max\{5p - 3, -7p + 4\} = -\frac{1}{12} \text{ (again with } p = \frac{7}{12}\text{)}$$

- **Example 2:** Collaborative two-finger Morra
  - Two Nash equilibria:
    - A and B both play 1 (value is 2)
    - A and B both play 2 (value is 4)

Summary

- **Two-player zero-sum games**
  - Von Neumann's minimax theorem
  - Multiple minimax strategies, single game value

- **Two-player non-zero-sum games**
  - Nash's existence theorem or fixed-point theorem
  - Multiple Nash equilibria, multiple game values

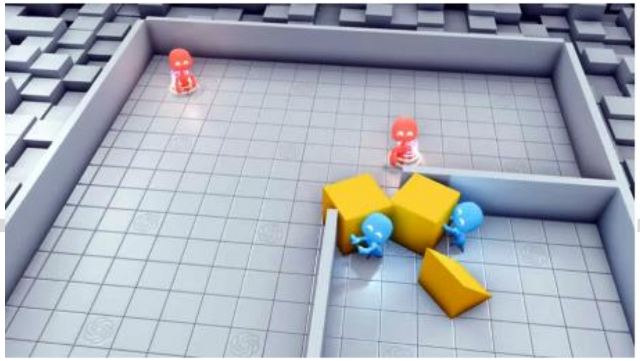
Other examples: AlphaGo

- Supervised learning: on human games
- Reinforcement learning: on self-play games
- Evaluation function: convolutional neural network (value network)
- Policy: convolutional neural network (policy network)
- Monte Carlo Tree Search: search/lookahead

Coordination games

- Hanabi: players need to signal to each other and coordinate in a decentralized fashion to collaboratively win. So, unlike most games, **you cannot see your own hand** — only your teammates can. You must rely on **limited communication** to figure out which cards to play

- Hide-and-Seek: OpenAI has developed multi-agent RL in which two teams of agents (hiders vs seekers) compete in a simulated physics-based world



Cooperative games

- A cooperative game is typically defined as a pair (N, v), where
  - $N = \{1,2,\ldots,n\}$ is the set of players
  - $v: 2^N \to \mathbb{R}$ is the characteristic function, which assigns a value to every coalition

- Applications:
  - **Resource allocation**: given a budget (e.g., one million dollars) and a list of public projects, allocate the budget on a subset of projects to maximize social welfare
  - **Robotics**: multi-agent cooperation

Solution concepts in cooperative games

- **Core**: A payoff vector $x = (x_1, x_2, \ldots, x_n)$ is in the core if $\sum_{i \in N} x_i = v(N)$, and $\sum_{i \in S} x_i \geq v(S), \forall S \subseteq N$
  - Example: in resource allocation games, this corresponds to solving a linear program

- **Shapley value**: ensures fair division https://en.wikipedia.org/wiki/Shapley_value