CS 407

Lecture 20: Learning in Games II

Best Response Dynamics

- Start with an arbitrary joint action a

- Repeat Forever:
  - Choose one player i to update strategy
    - Need to ensure each player will get chosen eventually
  - Compute $u_i(a', a_{-i})$ for each action a'
  - Set $a_i \in argmax_{a'} u_i(a', a_{-i})$

Fictitious Play (for 2 players)

- Initialize historical aggregate strategy $w_{-i} = (0, \ldots, 0)$

- Repeat Forever:
  - Update $w_{-i} \mathrel{+}= s_{-i}$ with the most recent opponent strategy
  - Compute average opponent strategy $\bar{w}_{-i} = w_{-i} / (\sum w_{-i})$
  - Compute $u_i(a', \bar{w}_{-i})$ for each action a'
  - Set $a_i \in argmax_{a'} u_i(a', \bar{w}_{-i})$

Smoothed Fictitious Play (for 2 players)

- Initialize historical aggregate strategy $w_{-i} = (0, \ldots, 0)$

- Repeat Forever:
  - Update w_{-i} += s_{-i} with the most recent opponent strategy
  - Compute average opponent strategy $\bar{w}_{-i} = w_{-i}/(\sum w_{-i})$
  - Compute $u_i(a', \bar{w}_{-i})$ for each action a'
  - Set $s_i(a_j) = \frac{\exp(u_i(a_j, \bar{w}_{-i})/\gamma)}{\sum_{a'} \exp(u_i(a', \bar{w}_{-i})/\gamma)}$
    - Equvalently choose s_i to maximize $u_i(s_i, \bar{w}_{-i}) - \gamma \sum_{a'} s_i(a') \log s_i(a')$

The prisoner’s dilemma

	Cooperate	Defect
Cooperate	-1,-1	-9,0
Defect	0,-9	-6,-6

Regret

- How much happier would I be if I played a_i instead of s_i?

$$u_i(a_i, s_{-i}) - u_i(s_i, s_{-i})$$

$$= u_i\ (a_i, s_{-i}) - \sum_{a'} s_i(a') u_i(a', s_{-i})$$

- Average Regret:

$$\frac{1}{T}\sum_{t=1}^{T}\left(u_i\left(a_i, s_{-i}^t\right) - u_i\left(s_i^t, s_{-i}^t\right)\right)$$

Regret of SFP

Theorem: Let $\epsilon > 0$ be given. Then for all a $\in A_i, \gamma < \gamma_0(\epsilon)$ if a player uses smoothed fictitious play with parameter γ then with probability 1 *regardless of the strategies of the other player*:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(u_i\left(a, s_{-i}^t\right) - u_i\left(a_i^t, s_{-i}^t\right) \right) < \epsilon$$

Regret Matching

- Initialize regret sums $R = (0, \ldots, 0)$

- Repeat Forever:
  - Update $R(a') \mathrel{+}= u_i(a', s\_i) - u_i(s_i, s\_i)$ with the most recent opponent strategy s_i and your last strategy s_i
  - Set $\mathrm{s_i}(a_j) = \frac{\max(R(a_j),0)}{\sum_{a'} \max(R(a'),0)}$
    - If denominator 0, set to uniform random

The prisoner’s dilemma

	Cooperate	Defect
Cooperate	-1,-1	-9,0
Defect	0,-9	-6,-6

Coordination Game

	Left	Right
Up	1,1	0,0
Down	0,0	1,1

Regret of Regret Matching

Theorem: If a player uses regret matching then with probability 1 *regardless of the strategies of the other player*:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(u_i\left(a_i, s_{-i}^t\right) - u_i\left(a_i^t, s_{-i}^t\right) \right) = 0$$

Convergence

- If I always do what smooth fictitious play / regret matching tells me to do, I will do (almost) as well as always choosing the best (fixed) action.

$$\sum_{a_1 \in A_1, a_2 \in A_2} p(a_1, a_2) u_1(a_1, a_2) \geq \sum_{a_1 \in A_1, a_2 \in A_2} p(a_1, a_2) u_1(a_1', a_2)$$

- p is a coarse correlated equilibrium (CCE) if both players are best responding in this sense

Convergence of Regret Minimization

Theorem: If all players use regret minimization algorithms such as smoothed fictitious play or regret matching their empirical joint strategy converges to a coarse correlated equilibrium.

(Approximately in the case of smoothed fictitious play)

Bach or Stravinsky

	Bach	Stravinsky
Bach	10,5	0,0
Stravinsky	0,0	5,10

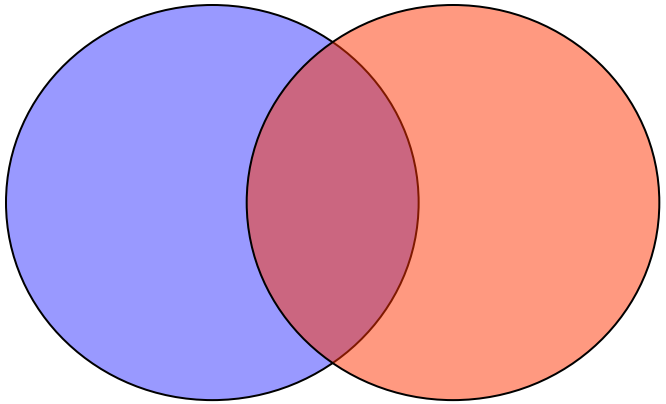
3rd Nash Equilibrium: $\left(\frac{2}{3},\frac{1}{3}\right),\left(\frac{1}{3},\frac{2}{3}\right)$

Correlated strategies

- Let $N = \{1,2\}$ for simplicity
- A mediator chooses a pair of actions (a_1, a_2) according to a joint distribution p over A
- Reveals a_1 to player 1 and a_2 to player 2

Conditioning

- Given two random variable X,Y with joint distribution p(x,y)
- If I learn that X = x, that tells me something about Y
- $P[Y = y \mid X = x] = \frac{\Pr[x \wedge y]}{\Pr[x]} = \frac{p(x,y)}{\sum_{y'} p(x,y')}$



Correlated strategies

- Let $N = \{1,2\}$ for simplicity
- A mediator chooses a pair of actions (a_1, a_2) according to a joint distribution p over A
- Reveals a_1 to player 1 and a_2 to player 2
- When player 1 gets $a_1 \in A_1$, he knows that the distribution over strategies of 2 is

$$\Pr[a_2|a_1] = \frac{\Pr[a_1 \wedge a_2]}{\Pr[a_1]} = \frac{p(a_1, a_2)}{\sum_{a_2' \in A_2} p(a_1, a_2')}$$

Correlated equilibrium

- Player 1 is best responding if for all $a_1' \in A_1$

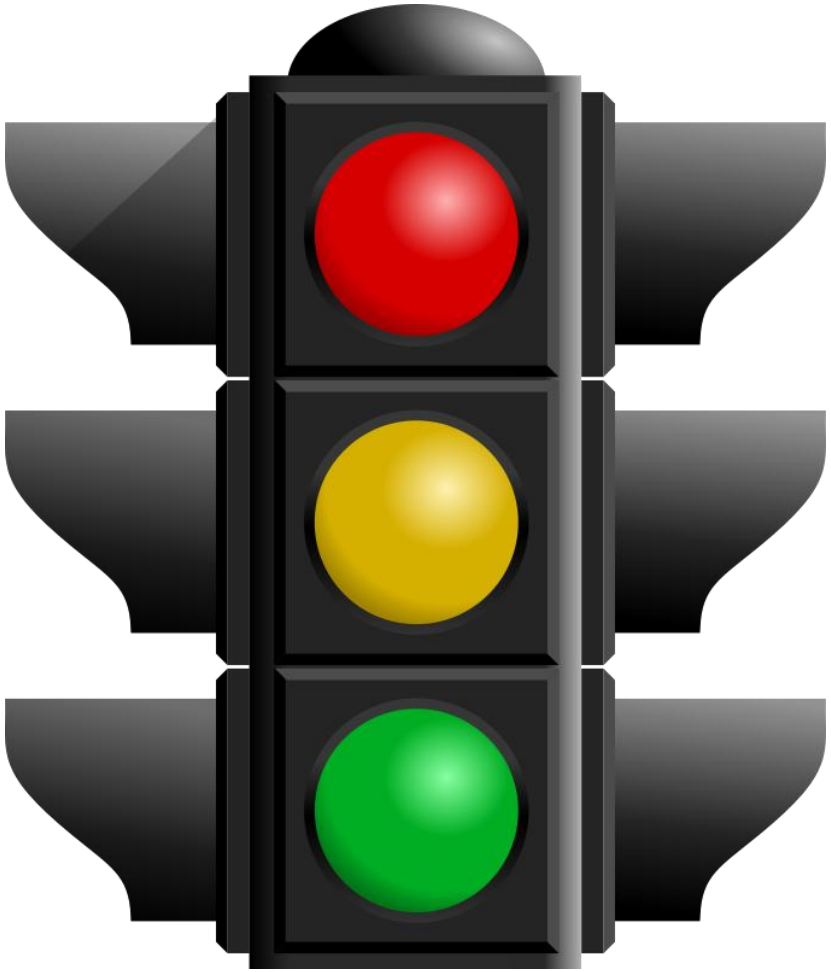
$$\sum_{a_2 \in A_2} \Pr[a_2|a_1]\, u_1(a_1, a_2) \geq \sum_{a_2 \in A_2} \Pr[a_2|a_1]\, u_1(a_1', a_2)$$

- Equivalently,

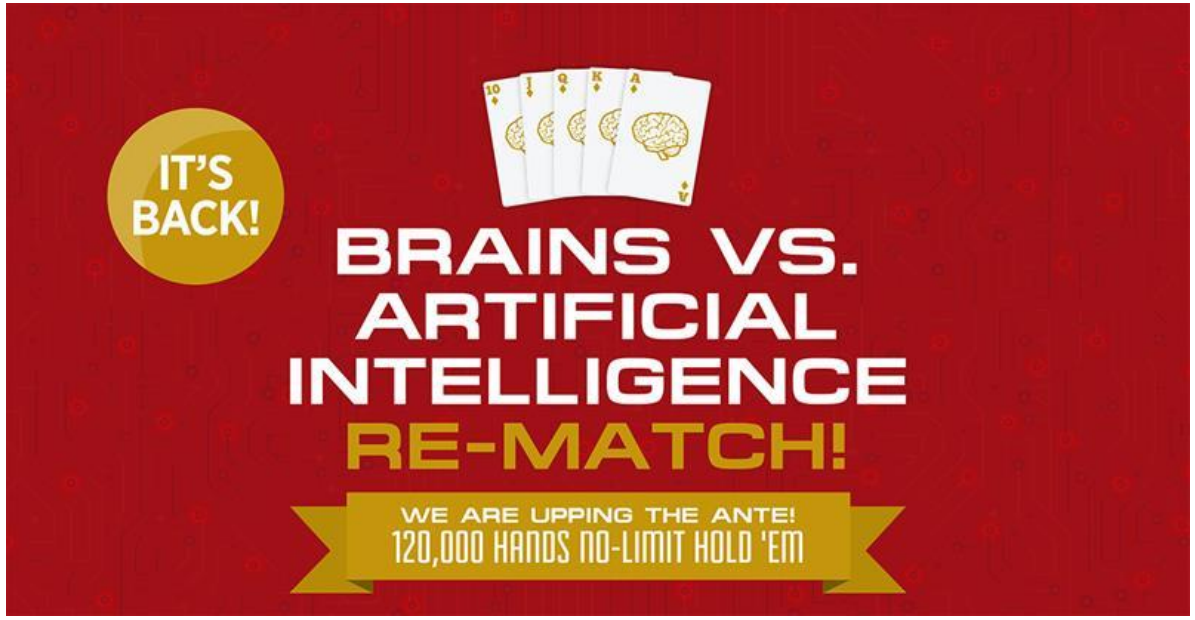
$$\sum_{a_2 \in A_2} p(a_1, a_2) u_1(a_1, a_2) \geq \sum_{a_2 \in A_2} p(a_1, a_2) u_1(a_1', a_2)$$

- p is a correlated equilibrium (CE) if both players are best responding

Correlated Equilibria You May Have Seen



Regret Minimization in Poker!



January 2017

DeepStack

December 2016

Key Takeaways

- Learning Algorithms:
  - Smoothed Fictitious Play
  - Regret Matching
- Other important ideas:
  - Regret Minimization
  - (Coarse) Correlated Equilibrium