Online learning in repeated matrix games

Yoav Freund

February 24, 2020

Based on “Adaptive Game Playing Using Multiplicative Weights” Freund and Schapire.

Outline

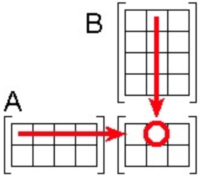
Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Row player gains $\mathbf{M}(i, j) \in [0, 1]$
- Column player looses $\mathbf{M}(i, j)$
- Game repeated many times.

Pure vs. mixed strategies

- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.
- Row player chooses dist. over rows $\mathbf{P}$
- Column player chooses dist. over columns $\mathbf{Q}$
- Row player gains $\mathbf{M}(\mathbf{P}, \mathbf{Q})$.
- Column player looses $\mathbf{M}(\mathbf{P}, \mathbf{Q})$.

Mixed strategies in matrix notation



$$(A \times B)_{12} = \sum_{r=1}^{4} a_{1r} b_{r2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} + a_{14}b_{42}$$

- **Q** is a column vector. $\mathbf{P}^T$ is a row vector.
- $\mathbf{M}(\mathbf{P}, \mathbf{Q}) = \mathbf{P}^T\mathbf{M}\mathbf{Q} = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{P}(i)\mathbf{M}(i,j)\mathbf{Q}(j)$

The minmax Theorem

When using pure strategies, second player has an advantage.

John von Neumann, 1928.

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) = \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

In words:

- for pure strategies, choosing second can be better.
- for mixed strategies, choosing second gives no advantage.
- There are min-max optimal mixed Strategies: $\mathbf{P}^*, \mathbf{Q}^*$
- $M(\mathbf{P}^*, \mathbf{Q}^*)$ is the value of the game.

Online Learning as matrix game

- Row = action
- Column = iteration.
- Player chooses mixed strategy $\mathbf{P}_t$
- adversary chooses pure strategy
  $\mathbf{Q}_t = \langle 0, \cdots, 0, 1, 0, \cdots, 0 \rangle$ the 1 is at position t
- Goal - minimize regret: $\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) - \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}^*, \mathbf{Q}_t)$

	$t = 1$	$t = 2$	...
expert1	0	1	...
expert2	0.2	0.1	...
expert3	0.5	0.2	...
...	...	...	...
Master	0.35	0.13	...

Boosting as a matrix game (1)

- Row = example (x, y)
- Column = Weak Rule h_t
- Matrix entry for $(x, y), h_t$ is 0 if $h_t(x) = y$, 1 $h_t(x) \neq y$

	h_1	h_2	...
*example*1	0	1	...
*example*2	1	0	...
*example*3	0	0	...
...	...	...	...

Boosting as a matrix game (2)

- Boosting assumption: for any distribution over examples, there exists a weak rule with weighted error $< 1/2$
- In game terms: For any mixed strategy of the row player $\mathbf{P}$, there is a pure strategy for column player $\mathbf{Q} = \langle 0, \cdots, 0, 1, 0, \cdots, 0 \rangle$ such that $M(\mathbf{P}, \mathbf{Q}) < 1/2)$
- From Min-Max theorem: There exists a column mixed strategy (a distribution over weak rules), that has expected value larger than zero for any row pure strategy (= any example).
- The weighted majority vote over the weak rule is **always** correct.

Adaboost as a repeated matrix game

- Booster chooses distribution over examples = mixed strategy over rows $\mathbf{P}_t$
- adversary chooses weak rule $\mathbf{Q}_t = \langle 0, \cdots, 0, 1, 0, \cdots, 0 \rangle$ the 1 is at position t
- **Goal 1:** produce a weighted majority rule that is highly accurate.
- **Goal 2:** Find a "hard" distribution over the training examples.

	h_1	h_2	...
*example*1	0	1	...
*example*2	1	0	...
*example*3	0	0	...
...	...	...	...

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.
- If all sides use learning, then game will converge to minmax equilibrium.
- If opponent is not optimally adversarial (limited by knowledge, computationa power...) then learning gives better performance than min-max.
- Our goal is to minimize regret.

Fictitious play

- also called “Follow the leader”
- Choose the best action with respect to the sum of past loss vectors.
- Might not converge to optimal mixed strategy.
- Consider playing the matching coins game against an adversary that alternates HTHTHTHTHT
- If #H > #T the next element is T
- If #T > #H the next element is H
- follow the leader makes an error on each iteration.

Randomized Fictitious play

- Also called 'Follow the perturbed leader'
- Choose the best action with respect to the sum of past loss vectors plus noise.
- Adding noise allows us to choose responses that are slightly worse than best response.
- Hannan 1957 Randomized ficticus play converges to regret minimizing strategy.
- regret is $O(1/\sqrt{n})$ where n is number of actions.

The basic algorithm

- Choose an initial distribution $\mathbf{P}_1$
- $$\mathbf{P}_{t+1}(i) = \mathbf{P}_t(i)\frac{e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}}{Z_t}$$
- Where $Z_t = \sum_{i=1}^{n} \mathbf{P}_t(i) e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}$
- $\eta > 0$ is the learning rate.

Generalized regret bound

- Regret relative to the best *pure strategy* i

$$\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t,\mathbf{Q}_t) \le \left(\frac{1}{1-e^{-\eta}}\right) \min_i \left[\eta \sum_{t=1}^{T}\mathbf{M}(i,\mathbf{Q}_t) - \ln \mathbf{P}_1(i)\right]$$

- regret with respect the the best *mixed strategy* $\mathbf{P}$:

$$\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t,\mathbf{Q}_t) \le \left(\frac{1}{1-e^{-\eta}}\right) \min_{\mathbf{P}} \left[\eta \sum_{t=1}^{T}\mathbf{M}(\mathbf{P},\mathbf{Q}_t) + \mathrm{RE}\left(\mathbf{P} \parallel \mathbf{P}_1\right)\right]$$

- Where

$$\mathrm{RE}\left(\mathbf{P} \parallel \mathbf{Q}\right) \doteq \sum_{i=1}^{n}\mathbf{P}(i)\ln\frac{\mathbf{P}(i)}{\mathbf{Q}(i)}$$

Main Theorem

- For any game matrix **M**.
- Any sequence of mixed strat. $\mathbf{Q}_1, \ldots, \mathbf{Q}_T$
- The sequence $\mathbf{P}_1, \ldots, \mathbf{P}_T$ produced by basic alg using $\eta > 0$ satisfies

$$\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \left(\frac{1}{1 - e^{-\eta}}\right) \min_{\mathbf{P}} \left[\eta \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + \mathrm{RE}\left(\mathbf{P} \parallel \mathbf{P}_1\right)\right]$$

Corollary

- Setting $\eta = \ln\left(1 + \sqrt{\frac{2\ln n}{T}}\right)$
- the average per-trial loss is

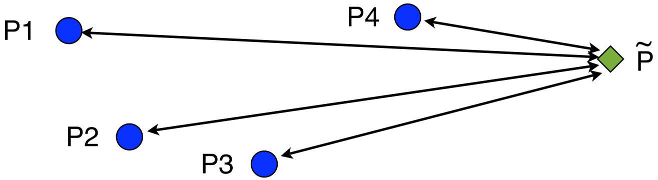
$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + \Delta_{T,n}$$

- Where

$$\Delta_{T,n} = \sqrt{\frac{2\ln n}{T}} + \frac{\ln n}{T} = O\left(\sqrt{\frac{\ln n}{T}}\right).$$

Visual intuition

- **Hedge**(η) : **If** $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \gg \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t)$ **then:** distance between $\mathbf{P}_{t+1}$ and $\tilde{\mathbf{P}}$ smaller than distance between $\mathbf{P}_t$ and $\tilde{\mathbf{P}}$
- $\mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_t\right) \leq \eta\mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$



The minmax Theorem

John von Neumann, 1928.

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) = \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

In words: for mixed strategies, choosing second gives no advantage.

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$
Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$ and $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$

$$
\begin{aligned}
\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} &\leq \max_{\mathbf{Q}} \overline{\mathbf{P}}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \\
&= \max_{\mathbf{Q}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \quad \text{by definition of } \overline{\mathbf{P}} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \max_{\mathbf{Q}} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}
\end{aligned}
$$

Proving minmax Theorem using online learning (2)

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\le \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t + \Delta_{T,n} && \text{by the Corollary} \\
&= \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\overline{\mathbf{Q}} + \Delta_{T,n} && \text{by definition of } \overline{\mathbf{Q}} \\
&\le \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} + \Delta_{T,n}.
\end{aligned}
$$

but $\Delta_{T,n}$ can be set arbitrarily small.

Solving a game

- to solve a game is to find the min-max mixed strategies $\mathbf{P}, \mathbf{Q}$
- Suppose that **Hedge**$(\eta)$is playing $\mathbf{P}_1, \mathbf{P}_2,$ against a worst case adversary that playes second: adversary that plays $\mathbf{Q}_1, \mathbf{Q}_2, \ldots$ such that $\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$.
- Without loss of generality $\mathbf{Q}_t$ is a pure strategy (prob. 1 on a single action).
- Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$, $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$

Using average distributions

- Von Neumann Min/Max Thm:
  $v \doteq \min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) = \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$
- Fixing T and letting $\eta = \ln\left(1 + \sqrt{\frac{2 \ln n}{T}}\right)$
- Two immediate corrolaries of the proof of the min/max Thm:

$$\max_{\mathbf{Q}} \mathbf{M}(\overline{\mathbf{P}}, \mathbf{Q}) \leq v + \Delta_{T,n}. \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \overline{\mathbf{Q}}) \geq v - \Delta_{T,n}$$

Using the final row distribution vMW

- Can we make the row distribution converge?
- Suppose we have an upper bound on the value of the game $u \geq v$
- **Good Enough:** If $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq u$ the row player does nothing $\mathbf{P}_{t+1} = \mathbf{P}_t$
- **Learn:** If $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) > u$ set

$$\eta = \ln \frac{(1-u)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)}{u(1-\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t))} \; .$$

Bound for vMW

- Let $\tilde{\mathbf{P}}$ be any mixed strategy for the rows such that $\max_{\mathbf{Q}} \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}) \leq u$
- Then on any iteration of algorithm vMW in which $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$ the relative entropy between $\tilde{\mathbf{P}}$ and $\mathbf{P}_{t+1}$ satisfies

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_{t+1}\right) \leq \mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_t\right) - \mathrm{RE}\left(u \parallel \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) .$$