Game-Theoretic Learning

Amy Greenwald
with David Gondek, Amir Jafari, and Casey Marks
Brown University

ICML Tutorial II

July 4, 2004

Overview

1. Introduction to Game Theory
2. Regret Matching Learning Algorithms
   - Regret Matching Learns Equilibria
3. Machine Learning Applications

Regret Matching Learning Algorithms

1. Regret Variations
   - No Φ-Regret Learning
   - External, Internal, and Swap Regret
2. Sufficient Conditions for No Φ-Regret Learning
   - Blackwell's Approachability Theorem
   - Gordon's Gradient Descent Theorem
     - Potential Function Argument
3. Expected and Observed Regret Matching Algorithms
   - Polynomial and Exponential Potential Functions
   - External, Internal, and Swap Regret
4. No Φ-Regret Learning Converges to Φ-Equilibria

   So Φ-Regret Matching Learns Φ-Equilibria

Regret Variations

Φ-Regret

1. External Regret
   - e.g., Freund and Schapire 97
2. Internal Regret
   - Foster and Vohra 99
3. Swap Regret
   - Blum and Mansour 04

Greenwald and Jafari 03

Road Map

Regret Vector

- External Regret $\leq$ Swap Regret
- Internal Regret $\leq$ Swap Regret

Regret Matrix

- External Regret $\leq (n-1)$ Internal Regret
- Swap Regret $\leq n$ Internal Regret

Single Agent Learning Model

- set of actions $N = \{1, \ldots, n\}$
- for all times t,
  - mixed action vector $q^t \in Q = \{q \in \mathbb{R}^n | \sum_i q_i = 1 \ \& \ q_i \geq 0,$
  - pure action vector $a^t = e_i$ for some pure action i
  - arbitrary reward vector $r^t = (r_1, \ldots, r_n)$

A learning algorithm $\mathcal{A}$ is a sequence of functions q^t : History^{t-1}
where a History is a sequence of action-reward pairs $(a^1, r^1), (a^2$

Pure Transformations

$$\mathcal{F}_{\mathsf{SWAP}} = \{F : N \rightarrow N\}$$
$$= \text{the set of all pure transformations}$$

$\mathcal{F}_{\mathsf{EXT}} = \{F^j \in \mathcal{F}_{\mathsf{SWAP}} | j \in N\}$, where $F^j(k) = j$

$\mathcal{F}_{\mathsf{INT}} = \{F^{ij} \in \mathcal{F}_{\mathsf{SWAP}} | ij \in N\}$, where $F^{ij}(k) = \begin{cases} j & \text{if } k \\ k & \text{oth} \end{cases}$

Mixed Transformations

$$\Phi_{\text{ALL}} = \{\phi : Q \to Q\}$$
$$= \text{the set of all mixed transformations}$$
$$= \text{the set of all row stochastic matrices}$$

$$\Phi_{\text{SWAP}} = \{\phi : Q \to Q | \phi \text{ deterministic}\} \subset \Phi_{\text{ALL}}$$

$$\Phi_{\text{EXT}} = \{\phi_j \in \Phi_{\text{SWAP}} | j \in N\}, \text{ where } e_k \phi_j = e_j$$

$$\Phi_{\text{INT}} = \{\phi_{ij} \in \Phi_{\text{SWAP}} | ij \in N\}, \text{ where } e_k \phi_{ij} = \begin{cases} e_j & \text{if } k \\ e_k & \text{othe} \end{cases}$$

Isomorphism

The operation of elements of $\mathcal{F}_{\mathsf{SWAP}}$ on $N \cong$
the operation of elements of Φ_{SWAP} on Q

$$\phi_{ij} = \delta_{F(i)=j}$$

Example If $|N| = 4$ and $F = \{1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 4, 4 \mapsto$

$$\phi = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$e_1\phi = e_2$, $e_2\phi = e_3$, $e_3\phi = e_4$, $e_4\phi = e_1$
In general, $e_k\phi = e_{F(k)}$, for all k

External Regret

$\mathcal{F}_{\text{EXT}} = \{F^j \in \mathcal{F}_{\text{SWAP}} | j \in N\}$, where $F^j(k) = j$
$\Phi_{\text{EXT}} = \{\phi^j \in \Phi_{\text{SWAP}} | j \in N\}$, where $e_k \phi^j = e_j$

Example If $|N| = 4$, then

$$\phi^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$e_1\phi^2 = e_2,\ e_2\phi^2 = e_2,\ e_3\phi^2 = e_2,\ e_4\phi^2 = e_2$

Internal Regret

$\mathcal{F}_{\text{INT}} = \{F^{ij} \in \mathcal{F}_{\text{SWAP}} | ij \in N\}$, where $F^{ij}(k) = \begin{cases} j \\ k \end{cases}$

$\Phi_{\text{INT}} = \{\phi^{ij} \in \Phi_{\text{SWAP}} | ij \in N\}$, where $e_k \phi^{ij} = \begin{cases} e_j & \text{if } k \\ e_k & \text{othe} \end{cases}$

Example If $|N| = 4$, then

$$\phi^{23} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$e_1 \phi^{23} = e_1$, $e_2 \phi^{23} = e_3$, $e_3 \phi^{23} = e_3$, $e_4 \phi^{23} = e_4$

Regret Vector $\rho \in \mathbb{R}^{\Phi}$

Observed Regret Vector $\rho_\phi(r,a) = r \cdot a\phi - r \cdot a$

Expected Regret Vector
$$\begin{aligned}\widehat{\rho}_\phi(r,q) &= \mathbb{E}[\rho_\phi(r,a) \mid a \sim q] \\ &= \rho_\phi(r, \mathbb{E}[a \mid a \sim q]) \\ &= r \cdot q\phi - r \cdot q\end{aligned}$$

Observed Regret $R(\Phi, T) \equiv R(\Phi, T, \{r\}, \{a\}) = \max_{\phi\in\Phi} \sum_{t=1}^{T} \rho_\phi$

Expected Regret $\widehat{R}(\Phi, T) \equiv \widehat{R}(\Phi, T, \{r\}, \{q\}) = \max_{\phi\in\Phi} \sum_{t=1}^{T} \widehat{\rho}_\phi$

Regret Definitions

	Observed	Expected
External	$R(\Phi_{\mathsf{EXT}}, T)$	$\hat{R}(\Phi_{\mathsf{EXT}}, T)$
Internal	$R(\Phi_{\mathsf{INT}}, T)$	$\hat{R}(\Phi_{\mathsf{INT}}, T)$
Swap	$R(\Phi_{\mathsf{SWAP}}, T)$	$\hat{R}(\Phi_{\mathsf{SWAP}}, T)$
All	$R(\Phi_{\mathsf{ALL}}, T)$	$\hat{R}(\Phi_{\mathsf{ALL}}, T)$

All Regret
An Upper Bound

Since $\Phi_{\mathsf{SWAP}} \subseteq \Phi_{\mathsf{ALL}}$,
swap regret is are bounded above by all regret:

$$\widehat{R}(\Phi_{\mathsf{SWAP}}, T) \;\leq\; \widehat{R}(\Phi_{\mathsf{ALL}}, T)$$

Lemma $\widehat{R}(\Phi_{\mathsf{SWAP}}, T) \geq \widehat{R}(\Phi_{\mathsf{ALL}}, T)$
Corollary $\widehat{R}(\Phi_{\mathsf{SWAP}}, T) = \widehat{R}(\Phi_{\mathsf{ALL}}, T)$

$$
\begin{aligned}
\widehat{R}(\Phi_{\mathsf{ALL}}, T) &= \max_{\phi \in \Phi_{\mathsf{ALL}}} \sum_{t=1}^{T} \rho_\phi(r^t, q^t) \\
&= \sum_{t=1}^{T} \rho_{\phi^*}(r^t, q^t) \\
&= \sum_{t=1}^{T} r^t \cdot q^t \phi^* - r^t \cdot q^t \\
&= \sum_{t=1}^{T} r^t \cdot q^t \left(\sum_{\phi \in \Phi_{\mathsf{SWAP}}} \lambda_\phi \phi \right) - r^t \cdot q^t \\
&= \sum_{\phi \in \Phi_{\mathsf{SWAP}}} \lambda_\phi \left(\sum_{t=1}^{T} r^t \cdot q^t \phi - r^t \cdot q^t \right) \\
&= \sum_{\phi \in \Phi_{\mathsf{SWAP}}} \lambda_\phi \left(\sum_{t=1}^{T} \rho_\phi(r^t, q^t) \right) \\
&\leq \max_{\phi \in \Phi_{\mathsf{SWAP}}} \sum_{t=1}^{T} \rho_\phi(r^t, q^t) \quad = \quad \widehat{R}(\Phi_{\mathsf{SWAP}}, T)
\end{aligned}
$$

Swap Regret
An Upper Bound

Since $\Phi_{\text{EXT}}, \Phi_{\text{INT}} \subseteq \Phi_{\text{SWAP}}$,
external and internal regret are bounded above by sw

$$\begin{aligned} \hat{R}(\Phi_{\text{EXT}}, T) &\leq \hat{R}(\Phi_{\text{SWAP}}, T) \\ \hat{R}(\Phi_{\text{INT}}, T) &\leq \hat{R}(\Phi_{\text{SWAP}}, T) \end{aligned}$$

Internal Regret
An Upper Bound

Claim External and swap regret are bounded above $n-1$ and n times internal regret, respectively:

$$
\begin{aligned}
\widehat{R}(\Phi_{\mathsf{EXT}}, T) &\leq (n-1)\widehat{R}(\Phi_{\mathsf{INT}}, T) \\
\widehat{R}(\Phi_{\mathsf{SWAP}}, T) &\leq n\widehat{R}(\Phi_{\mathsf{INT}}, T)
\end{aligned}
$$

Regret Matrix

Observed Regret Matrix $m_{ij}^t \equiv a_i^t(r_j^t - r_i^t)$

Expected Regret Matrix

$$\begin{aligned}\widehat{m}_{ij}^t &\equiv \mathbb{E}[a_i^t(r_j^t - r_i^t) \mid a^t \sim q^t] \\ &= \mathbb{E}[a_i^t \mid a^t \sim q^t](r_j^t - r_i^t) \\ &= q_i^t(r_j^t - r_i^t)\end{aligned}$$

Cumulative Observed Regret Matrix $M^T = \sum_{t=1}^T m_{ij}^t$

Cumulative Expected Regret Matrix $\widehat{M}^T = \sum_{t=1}^T \widehat{m}_{ij}^t$

$$
\begin{aligned}
\widehat{R}(\Phi,T) &= \max_{\phi\in\Phi}\sum_{t=1}^{T}\widehat{\rho}_{\phi}(r^t,q^t) \\
&= \max_{\phi\in\Phi}\sum_{t=1}^{T} r^t\cdot q^t\phi - r^t\cdot q^t \\
&= \max_{\phi\in\Phi}\sum_{t=1}^{T}\left(\sum_{j\in N} r_j^t (q^t\phi)_j - \sum_{i\in N} r_i^t q_i^t\right) \\
&= \max_{\phi\in\Phi}\sum_{t=1}^{T}\left(\sum_{j\in N} r_j^t \sum_{i\in N} q_i^t\phi_{ij} - \sum_{i\in N} r_i^t q_i^t\right) \\
&= \max_{F\in\mathcal{F}}\sum_{t=1}^{T}\left(\sum_{j\in N} r_j^t \sum_{i\in N} q_i^t\delta_{F(i)=j} - \sum_{i\in N} r_i^t q_i^t\right) \\
&= \max_{F\in\mathcal{F}}\sum_{t=1}^{T}\left(\sum_{i\in N} r_{F(i)}^t q_i^t - \sum_{i\in N} r_i^t q_i^t\right) \\
&= \max_{F\in\mathcal{F}}\sum_{t=1}^{T}\sum_{i\in N} q_i^t (r_{F(i)}^t - r_i^t) \\
&= \max_{F\in\mathcal{F}}\sum_{i\in N}\widehat{M}_{iF(i)}^T
\end{aligned}
$$

External Regret

$$
\begin{aligned}
\widehat{R}(\Phi_{\mathsf{EXT}}, T) &= \max_{F \in \mathcal{F}_{\mathsf{EXT}}} \sum_{i \in N} \widehat{M}^T_{iF(i)} \\
&= \max_{j \in N} \sum_{i \in N} \widehat{M}^T_{iF^j(i)} \\
&= \max_{j \in N} \sum_{i \in N} \widehat{M}^T_{ij}
\end{aligned}
$$

Internal Regret

$$\begin{aligned}
\widehat{R}(\Phi_{\mathsf{INT}}, T) &= \max_{F \in \mathcal{F}_{\mathsf{INT}}} \sum_{k \in N} \widehat{M}^T_{kF(k)} \\
&= \max_{ij \in N} \sum_{k \in N} \widehat{M}^T_{kF^{ij}(k)} \\
&= \max_{ij \in N} \left(\widehat{M}^T_{ij} + \sum_{k \neq i} \widehat{M}^T_{kk} \right) \\
&= \max_{ij \in N} \widehat{M}^T_{ij}
\end{aligned}$$

Swap Regret

$$
\begin{aligned}
\hat{R}(\Phi_{\mathsf{SWAP}}, T) &= \max_{F \in \mathcal{F}_{\mathsf{SWAP}}} \sum_{i \in N} \widehat{M}^T_{iF(i)} \\
&= \max_{z_1 \in N} \ldots \max_{z_n \in N} \left(\widehat{M}^T_{1z_1} + \ldots + \widehat{M}^T_{nz_n} \right) \\
&= \max_{z_1 \in N} \widehat{M}^T_{1z_1} + \ldots + \max_{z_n \in N} \widehat{M}^T_{nz_n} \\
&= \sum_{i \in N} \max_{j \in N} \widehat{M}^T_{ij}
\end{aligned}
$$

Summary

$$
\begin{aligned}
\widehat{R}(\Phi_{\mathsf{INT}}, T) &= \max_{ij \in N} \widehat{M}_{ij}^T \\
\widehat{R}(\Phi_{\mathsf{EXT}}, T) &= \max_{j \in N} \sum_{i \in N} \widehat{M}_{ij}^T \\
\widehat{R}(\Phi_{\mathsf{SWAP}}, T) &= \sum_{i \in N} \max_{j \in N} \widehat{M}_{ij}^T
\end{aligned}
$$

External Regret Bound

If $X_{ii} = 0$, then

$$\sum_{i \in N} X_{ij} \leq (n-1) \max_i X_{ij}, \quad \forall j$$

Therefore

$$\begin{aligned}
\hat{R}(\Phi_{\mathsf{EXT}}, T) &= \max_{j \in N} \sum_{i \in N} \widehat{M}_{ij}^T \\
&\leq \max_{j \in N} \left((n-1) \max_{i \in N} \widehat{M}_{ij}^T \right) \\
&= (n-1) \max_{ij \in N} \widehat{M}_{ij}^T \\
&= (n-1) \hat{R}(\Phi_{\mathsf{INT}}, T)
\end{aligned}$$

Example

Action Sequence $123\ldots n$

Optimal Sequence $111\ldots 1$

n External Regret Sequences:

$\{111\ldots 1\}$, $\{222\ldots 2\}$, ..., $\{nnn\ldots n\}$

(Observed) External Regret is $n-1-0=n-1$.

n^2 Internal Regret Sequences:

$\{123\ldots n\}$, $\{113\ldots n\}$, $\{121\ldots n\}$, ..., $\{123\ldots n-2, n-1, 1\}$, ..

$\{n23\ldots n\}$, $\{1n3\ldots n\}$, $\{12n\ldots n\}$, ..., $\{123\ldots n-2, n-1, n\}$

(Observed) Internal Regret is $n-1-(n-2)=1$.

n^n Swap Regret Sequences:

(Observed) Swap Regret is $n-1-0=n-1$.

Example

Action Sequence $123\ldots n$

Optimal Sequence $111\ldots 1$

Cumulative (Observed) Regret Matrix
$$\begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\begin{aligned} R(\Phi_{\mathsf{INT}}, n) &= \max_{ij \in N} M_{ij}^n = 1 \\ R(\Phi_{\mathsf{EXT}}, n) &= \max_{j \in N} \sum_{i \in N} M_{ij}^n = n - 1 \\ R(\Phi_{\mathsf{SWAP}}, n) &= \sum_{i \in N} \max_{j \in N} M_{ij}^n = n - 1 \end{aligned}$$

Swap Regret Bound

Observe

$$\sum_{i\in N} \max_{j\in N} X_{ij} \leq n \max_{ij\in N} X_{ij}, \quad \forall j$$

Therefore

$$\begin{aligned}\widehat{R}(\Phi_{\mathsf{SWAP}}, T) &= \sum_{i\in N} \max_{j\in N} \widehat{M}_{ij}^{T} \\ &\leq n \max_{ij\in N} \widehat{M}_{ij}^{T} \\ &= n\widehat{R}(\Phi_{\mathsf{INT}}, T)\end{aligned}$$

Example

Action Sequence $123\dots n$

Optimal Sequence $234\dots 1$

n External Regret Sequences:

$\{111\dots 1\}$, $\{222\dots 2\}$, ..., $\{nnn\dots n\}$

(Observed) External Regret is $n-(n-1)=1$.

n^2 Internal Regret Sequences:

$\{123\dots n\}$, $\{113\dots n\}$, $\{121\dots n\}$, ..., $\{123\dots n-2, n-1, 1\}$

$\{223\dots n\}$, $\{123\dots n\}$, $\{122\dots n\}$, ..., $\{123\dots n-2, n-1, 2\}$, ..

(Observed) Internal Regret is $n-(n-1)=1$.

n^n Swap Regret Sequences:

(Observed) Swap Regret is $n-0=n$.

Example

Action Sequence $123 \ldots n$

Optimal Sequence $234 \ldots 1$

Cumulative (Observed) Regret Matrix $\begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$

$$
\begin{aligned}
R(\Phi_{\mathsf{INT}}, n) &= \max_{ij \in N} M^n_{ij} = 1 \\
R(\Phi_{\mathsf{EXT}}, n) &= \max_{j \in N} \sum_{i \in N} M^n_{ij} = 1 \\
R(\Phi_{\mathsf{SWAP}}, n) &= \sum_{i \in N} \max_{j \in N} M^n_{ij} = n
\end{aligned}
$$

Regret Bounds

External Regret $\leq$ Swap Regret

Internal Regret $\leq$ Swap Regret

External Regret $\leq (n-1)$ Internal Regret

Swap Regret $\leq n$ Internal Regret

No Swap Regret $\Rightarrow$ No External Regret

No Swap Regret $\Rightarrow$ No Internal Regret

No Internal Regret $\Rightarrow$ No External Regret

No Internal Regret $\Rightarrow$ No Swap Regret

No Φ-Regret Learning

A learning algorithm $\mathcal{A} = q^1, q^2, \ldots$ exhibits no expected Φ-regret
iff

$$\limsup_{T \to \infty} \frac{\widehat{R}(\Phi, T, \{r\}, \{q\})}{T} \leq 0$$

A learning algorithm $\mathcal{A} = q^1, q^2, \ldots$ exhibits no observed Φ-regret
iff

$$\limsup_{T \to \infty} \mathbb{E}\left[\frac{R(\Phi, T, \{r\}, \{a\})}{T} \mid \{a \sim q\}\right] \leq 0$$

No Observed Regret, i.e. ⇒ No Expected Re

$$
\begin{aligned}
\widehat{R}(\Phi, T, \{r\}, \{q\}) &= \max_{\phi\in\Phi} \sum_{t=1}^{T} \widehat{\rho}(r^t, q^t) \\
&= \max_{\phi\in\Phi} \sum_{t=1}^{T} \mathbb{E}\left[\rho(r^t, a^t) | a^t \sim q^t\right] \\
&= \max_{\phi\in\Phi} \sum_{t=1}^{T} \sum_{a^1,\ldots,a^t} q^t(a^1, \ldots, a^t) \rho(r^t, a^t) \\
&= \max_{\phi\in\Phi} \sum_{a^1,\ldots,a^T} q^T(a^1, \ldots, a^T) \sum_{t=1}^{T} \rho(r^t, a^t) \\
&= \max_{\phi\in\Phi} \mathbb{E}\left[\sum_{t=1}^{T} \rho(r^t, a^t) | \{a \sim q\}\right] \\
&\leq \mathbb{E}\left[\max_{\phi\in\Phi} \sum_{t=1}^{T} \rho(r^t, a^t) | \{a \sim q\}\right] \\
&= \mathbb{E}\left[R(\Phi, T, \{r\}, \{a\} | \{a \sim q\}\right]
\end{aligned}
$$

Summary of Part II

$$
\begin{array}{ccc}
 & \text{No Expected External Regret} & \\
 & \Uparrow & \\
 & \text{No Observed External Regret, i.e.} & \\
 & \Uparrow & \\
\text{No Observed Swap Regret, i.e.} & \Leftrightarrow & \text{No Observed Internal Regret} \\
\Downarrow & & \Downarrow \\
\text{No Expected Swap Regret} & \Leftrightarrow & \text{No Expected Internal Regret} \\
 & \Downarrow & \\
 & \text{No Expected External Regret} &
\end{array}
$$