DS 320 Algorithms for Data Science **Lecture #21 Worksheet**
Spring 2022 Prof. Kira Goldner

Zero-Sum Games and the Minimax Theorem

	Rock	Paper	Scissors
Rock	0	-1	1
Paper	1	0	-1
Scissors	-1	1	0

The Minimax Theorem

Theorem 1 (Minimax Theorem)**.** *For every two-player zero-sum game* $\mathbf{A}$*,*

$$\max_{\mathbf{x}} \left(\min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) = \min_{\mathbf{y}} \left(\max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right). \tag{1}$$

From LP Duality to Minimax

$$\max_{\mathbf{x}} \left(\min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \mathbf{x}^T \mathbf{A} \mathbf{e}_j \right) \tag{2}$$

$$= \max_{\mathbf{x}} \left(\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i \right) \tag{3}$$

$$\max v$$

subject to

$$\begin{aligned} v - \sum_{i=1}^{m} a_{ij} x_i \leq 0 \quad & \text{for all } j = 1, \ldots, n \\ \sum_{i=1}^{m} x_i = 1 & \\ x_1, \ldots, x_m \geq 0 \quad & \text{and} \quad v \in \mathbb{R}. \end{aligned}$$

$$\min w$$

subject to

$$\begin{aligned} w - \sum_{j=1}^{n} a_{ij} y_j &\geq 0 \quad \text{for all } i = 1, \ldots, m \\ \sum_{j=1}^{n} y_j &= 1 \\ y_1, \ldots, y_n &\geq 0 \quad \text{and} \quad w \in \mathbb{R}. \end{aligned}$$

Online Learning and the Multiplicative Weights Algorithm

An Online Problem

1. The input arrives "one piece at a time."
2. An algorithm makes an irrevocable decision each time it receives a new piece of the input.

Online Decision-Making

At each time step $t = 1, 2, \ldots, T$:

a decision-maker picks a probability distribution $\mathbf{p}^t$ over her actions A

an adversary picks a reward vector $\mathbf{r}^t : A \rightarrow [-1, 1]$

an action at is chosen according to the distribution $\mathbf{p}^t$, and the decision-maker receives reward $r^t(a^t)$

the decision-maker learns $\mathbf{r}^t$, the entire reward vector

The input arrives “one piece at a time.”

What should we compare to?

Definition 1 (Regret)**.** Fix reward vectors $\mathbf{r}^1, \ldots, \mathbf{r}^T$. The regret of the action sequence $a^1, \ldots, a^T$ is

$$\underbrace{\max_{i=1}^{N} \sum_{t=1}^{T} r_i^t}_{\text{best fixed action}} - \underbrace{\sum_{t=1}^{T} r^t(a^t)}_{\text{our algorithm}} . \tag{4}$$

No-Regret Algorithm Design Principles

1. Past performance of actions should guide which action is chosen at each time step, with the probability of choosing an action increasing in its cumulative reward. (Recall from Example 2.3 that we need a randomized algorithm to have any chance.)
2. The probability of choosing a poorly performing action should decrease at an exponential rate.