CSE599s, Spring 2012, Online Learning	Lecture 2 - 03/29/2012
The online optimization game	
Lecturer: Brendan McMahan or Ofer Dekel	*Scribe: Marzieh Nabi*

1 Announcement

Recently, journal of Foundations and Trends in Machine Learning published a paper by Shai Shalev-Shwartz on "Online Learning and Online Convex Optimization", [1], which can be considered as a good reference for this course.

2 Introduction

As a summary of the previous lecture, consider a finite class of experts called H. Consider two following cases

- if $\exists\ h^* \in H$ which is perfect, Halving algorithms claims $M \leq \log_2(|H|)$ where M is the number of mistakes
- if H is a class of linear classifiers and $\exists\ h^* \in H$ perfect with a margin $\gamma > 0$, Perceptron algorithm says $M \leq \left(\frac{\rho}{\gamma}\right)^2$.

The concern with these algorithms is that the **realizability** assumption is unrealistic. Now let us define a new game.

3 The online optimization game (between a player and an adversary)

For $t = 1, 2, \ldots$, here is the game

1. player chooses $w_t \in \mathcal{W}$ where $\mathcal{W} \subseteq \mathbb{R}^n$ is a feasible set of actions.
2. the adversary has infinite power and it chooses a loss function $f_t : \mathcal{W} \to \mathbb{R}$.
3. the player suffers the loss function $f_t(w_t)$
4. the player updates w_{t+1}

Let us start with the following definition

Definition 1. *The cumulative loss after T rounds of play is $\sum_{t=1}^{T} f_t(w_t)$.*

What is the goal?
The goal is to

- minimize the cumulative loss: the adversary can make the loss arbitrary large.
- we need to compare to a benchmark.

Definition 2. *The player's regret after T rounds is defined $\sum_{t=1}^{T} f_t(w_t) - \min\limits_{w \in \mathcal{W}} \sum_{t=1}^{T} f_t(w)$.*

In words, we compare our cumulative loss to the **"best" fixed** point in hindsight.

Goal:

- prove regret bounds $R(T)$-function that upper bounds the regret $\forall\ T,\ \forall f_1, f_2, \ldots$
- $R(T)$-sublinear, $R(T) = O(\sqrt{T}, T^{2/3}, \ldots)$
- later we will relax **"fixed"** in various ways
- $\mathcal{W}$ plays two roles
  1. set of actions for the player
  2. the "comparison" or the "competitor" class
- the adversary and the competitor class are two different entities

4 Online binary prediction

The online binary prediction is a special case of online optimization.

online binary prediction	**online optimization**
for $t = 1, 2, \ldots$ choose $h_t \in H$	for $t = 1, 2, \ldots$ choose $w_t \in \mathcal{W}$
the adversary chooses (x_t, y_t)	the adversary chooses loss $f_t(w_t)$
loss at time t is equal to $\mathbf{1}_{h_t(x_t) \neq y_t}$	we suffer the loss function $f_t(w_t)$

We can think of h_t as parametrized by w_t and the loss function $f_t(w)$, in the online optimization, is $f_t(w) = \mathbf{1}_{h_w(x_t) \neq y_t}$.

5 Online convex optimization

In this problem

1. $\mathcal{W}$ is a convex subset of $\mathbb{R}^n$
2. f_t is convex for all t

Following is couple examples

- **Online linear regression** $w_t \in \mathcal{W} = \mathbb{R}^n,\ x_t \in \mathbb{R}^n,\ y_t \in \mathbb{R}, f_t(w_t) = (w_t \cdot x_t - y_t)^2$
- deg-2 linear regression $(w_t, A_t) \in (\mathbb{R}^n \times \mathbb{R}^{n \times n}),\ x_t \in \mathbb{R}^n,\ y_t \in \mathbb{R}, f_t(w_t, A_t) = (w_t \cdot x_t + x_t^T A_t x_t - y_t)^2$
- Air-dropping supplies $\mathcal{W} = \mathbb{R}^2,\ x_t \in \mathbb{R}^2, f_t(w_t) = ||w_t - x_t||_2$ or other norms such as $f_t(w_t) = ||w_t - x_t||_1 = \sum_{j=1}^{n} |w_{tj} - x_{tj}|$ like the Manhattan police example
- online portfolio management
- online to Batch conversion: a way to use an online learning algorithm for an offline learning algorithm

6 Follow the leader (FTL) algorithm

The update strategy in this algorithm in this algorithm is as follows

$$
\begin{aligned}
w_{t+1} &= \underset{w \in \mathcal{W}}{arg\ \min} \sum_{s=1}^{t} f_s(w) \\
&= \underset{w \in \mathcal{W}}{arg\ \min}\ f_{1:t}(w)
\end{aligned}
$$

where the notation $f_{1:t}(w) = \sum_{s=1}^{t} f_s(w)$ captures a batch of all data from time 1 to t.

Example 1 Consider $\mathcal{W} \in [-1, 1]$, $f_t(w) = g_t.w$ where $g_t \in [-1, 1]$. At time t, we take the action w_t and the adversary chooses g_t and we suffer the loss. Then, we have the following table of couple steps

t	w_t	g_t	loss	$g_{1:t}$
1	0	0.5	0	0.5
2	-1	-1	1	-0.5
3	1	1	1	0.5
4	-1	-1	1	-.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

For example $w_2 = \underset{w \in [-1,1]}{arg\ \min}\ 0.5w$. Therefore, the regret is of order T, $R(T) = O(T)$.

References

[1] Shai Shalev-Shwartz, "Online Learning and Online Convex Optimization", Foundations and Trends in Machine Learning, 2012.