[CS3958: Lecture 6] Online Learning, Convex Optimization

Instructor: Chihao Zhang, Scribed by Yulin Wang

November 9, 2022

The multi-armed bandit problem we met in the last lecture is a typical online-decision making problem. Today we will first examine a few other similar examples and introduce the framework of online learning.

1 An Online Number Guessing Game

Let us first play the following game. The game lasts for T rounds, and in each round $t = 1, 2, \dots, T$:

- The adversary picks a number $y_t \in [0, 1]$;
- You pick a number $x_t \in [0, 1]$ (without knowing y_t);
- The number y_t is revealed and you pay the cost $(x_t - y_t)^2$.

The goal is to minimize the cumulative cost $\sum_{t=1}^{T}(x_t - y_t)^2$. The strategy of the player depends on how the adversary picks the number y_t. There are two typical settings:

- **Stochastic Setting**: Each y_t is drawn from some fixed distribution $\mathcal{D}$;
- **Adversarial Setting**: The adversary can arbitrarily pick a legal y_t as if he knows the player's strategy.

1.1 The Stochastic Setting

Let us assume $y_t \sim \mathcal{D}$ for some fixed distribution $\mathcal{D}$ independently for each $t \in [T]$. Then the cumulative cost is a random variable and we want to minimize $\mathbf{E}\left[\sum_{t=1}^{T}(x_t - y_t)^2\right]$.

Since each round is independent, we only need to determine x_t to minimize $\mathbf{E}\left[(Y - x_t)^2\right]$ for $Y \sim \mathcal{D}$. This is equivalent to minimizing $-2\mathbf{E}[Y]\,x_t + x_t^2$ and the minimum is achieved at $x_t = \mathbf{E}[Y]$. Therefore, if we know the distribution $\mathcal{D}$ *in advance*, the optimal strategy is to pick $x_t = \mathbf{E}[Y] =: x^*$ in each round.

What if $\mathcal{D}$ is not known? Since any strategy can not beat the optimal choice of $x^* = \mathbf{E}[Y]$, it is natural to benchmark the cost using the following quantity, the regret $R(T)$, meaning *the regret of not playing* x^*:

$$R(T) = \sum_{t=1}^{T}(x_t - y_t)^2 - \min_{x\in[0,1]}\sum_{t=1}^{T}(x - y_t)^2 = \sum_{t=1}^{T}(x_t - y_t)^2 - \sum_{t=1}^{T}(x^* - y_t)^2.$$

1.2 The Adversarial Setting

In the adversarial setting, the number y_t can be arbitrarily picked and the player needs to determine x_t without knowing y_t. Therefore, in order to emphasis the adversarial nature, in the description below, we may assume that y_t is picked after x_t is determined. Similar to the stochastic setting, we are also interested in the regret

$$R(T) = \max_{y_1, y_2, \dots, y_T} \left(\sum_{t=1}^{T} (x_t - y_t)^2 - \min_{x \in [0,1]} \sum_{t=1}^{T} (x - y_t)^2 \right).$$

One needs to justify the use of regret as a measure of the performance of the algorithm in the adversarial setting as it is clear from the definition that $R(T)$ might be negative. The two points regarding this would be much more clear later.

- In a canonical example of online learning, the problem of learning from expert advice (we will meet the problem in the next lecture), the regret means the difference of the performance between your strategy and the best expert in hindsight.
- The quantity is closely related to the offline optimization version of the problem.

1.3 *The* Follow-The-Leader *Algorithm for the Game*

We now describe a good strategy to play the game. We again begin with solving the offline version of the game, namely $\min \sum_{t=1}^{T} (x - y_t)^2$. Taking derivative on x, we can obtain

$$x^* = \arg\min_{x \in [0,1]} \sum_{t=1}^{T} (x - y_t)^2 = \frac{1}{T} \sum_{t=1}^{T} y_t.$$

That is, the optimal choice of x is the mean of $y_1, y_2, \dots, y_T$.

On the otherhand, if the game is played online, the player only knows $y_1, \dots, y_{t-1}$ when it is up for him to determine x_t. Therefore, a possible good choice is to choose x_t as the mean of $y_1, \dots, y_{t-1}$ since the number $\frac{1}{t-1} \sum_{s=1}^{t-1} y_s$ is the offline optimal if the game is only played for $t - 1$ rounds. Our strategy is just *following the leader in the last round.* In the following, for every $t = 1, \dots, T$, we use x_t^* to denote $\frac{1}{t} \sum_{s=1}^{t} y_s$, the optimal solution up to round t.

Theorem 1 *Let $y_t \in [-1, 1]$ for $t = 1, \dots, T$ be an arbitrary sequence of numbers. Assume we choose $x_t = x_{t-1}^* = \frac{1}{t-1} \sum_{s=1}^{t-1} y_s$. Then*

$$R(T) = \sum_{t=1}^{T} (x_t - y_t)^2 - \min_{x \in [0,1]} \sum_{t=1}^{T} (x - y_t)^2 \leq 2H_T = \Theta(\log T),$$

where $H_T = \sum_{t=1}^{T} \frac{1}{t}$ is the T-th harmonic number.

Proof. We first prove the inequality *the cost of non-adaptive global optimal is greater than the cost of adaptive optimals*:

$$\sum_{t=1}^{T}(x_T^* - y_t)^2 \geq \sum_{t=1}^{T}(x_t^* - y_t)^2. \tag{1}$$

We prove (1) by induction on T. The base case of $T = 1$ clearly holds. So we assume (1) holds for smaller T. Note that

$$\sum_{t=1}^{T}(x_T^* - y_t)^2 \geq \sum_{t=1}^{T}(x_t^* - y_t)^2 \iff \sum_{t=1}^{T-1}(x_T^* - y_t)^2 \geq \sum_{t=1}^{T-1}(x_t^* - y_t)^2.$$

The induction hypothesis then implies

$$\sum_{t=1}^{T-1}(x_t^* - y_t)^2 \leq \sum_{t=1}^{T-1}(x_{T-1}^* - y_t)^2 \leq \sum_{t=1}^{T-1}(x_T^* - y_t)^2.$$

With (1), we can bound the regret as

$$R(T) = \sum_{t=1}^{T}(x_t - y_t)^2 - \sum_{t=1}^{T}(x_T^* - y_t)^2 \leq \sum_{t=1}^{T}(x_t - y_t)^2 - \sum_{t=1}^{T}(x_t^* - y_t)^2.$$

For each $t = 2, \dots, T$, we have

$$\begin{aligned}
(x_t - y_t)^2 - (x_t^* - y_t)^2 &= (x_{t-1}^* + x_t^* - 2y_t)(x_{t-1}^* - x_t^*) \\
&\leq 2\left|x_{t-1}^* - x_t^*\right| \\
&= 2\left|\frac{1}{t-1}\sum_{s=1}^{t-1} y_s - \frac{1}{t}\sum_{s=1}^{t} y_s\right| \\
&= 2\left|\frac{1}{t(t-1)}\sum_{s=1}^{t-1} y_s - \frac{1}{t} y_t\right| \\
&\leq \frac{2}{t}.
\end{aligned}$$

Then $R(T) \leq \sum_{t=1}^{T} \frac{2}{t} = 2H_T$. □

2 Online Learning

We can generalize the setting of the above game, which yields the framework of online learning / online optimization. Let $V \subseteq \mathbb{R}^d$. We again play a game for T rounds, and for each $t = 1, 2, \dots, T$:

- You pick some $x_t \in V$;
- The adversary pick some $\ell_t : V \longrightarrow \mathbb{R}$;
- The function ℓ_t is revealed and you pay the cost $\ell_t(x_t)$.

We want to minimize the regret

$$R(T) = \sum_{t=1}^{T} \ell_t(x_t) - \min_{x \in V} \sum_{t=1}^{T} \ell_t(x).$$

In case the strategy is randomized, we turn to minimize $\mathbf{E}[R(T)]$. The *follow-the-leader (FTL)* strategy can also be applied to this general setting. That is, we always choose

$$x_t = \arg\min_{x} \sum_{s=1}^{t-1} \ell_t(x).$$

The analysis of FTL in the previous problem also applies to the general case. Astute reader may already discovered that the regret $R(T)$ relies on the *sensitity* of the *intermediate optimal solutions*: $\ell_t(x^*_{t-1}) - \ell_t(x^*_t)$. That is, how much the lost up to round t changes if one replaces the optimal solution by the optimal solution of $t-1$ rounds. We will see an example in which the optimal solutions greatly oscillates as the game proceeds, and therefore FTL is not a good strategy.

Example 1 (Failure of FTL) *Consider $V = [0, 1]$, $\ell_t(x) = a_t x$ where $a_1 = -0.5$ and for $t > 1$,*

$$a_t = \begin{cases} 1 & \text{if } t \text{ is even;} \\ -1 & \text{if } t \text{ is odd.} \end{cases}$$

Then FTL tells you one should choose $x_t = 1$ for even t and $x_t = -1$ for odd t when $t > 1$. However, the gap between the cost of this strategy and the cost of $x_t = 0$ for all t is linear in T.

Moreover, in order to apply FTL, one needs to solve the offline optimization problem $x_t = \arg\min_x \sum_{s=1}^{t-1} \ell_t(x)$, which is already non-trivial or even computational hard in some cases. This motivates us to first look at offline optimization problems, which seem to be easier.