CS170–Fall 2022 — Homework 9 Solutions

Deming Chen cdm@pku.edu.cn

Last Modified: January 8, 2023

Collaborators: NONE

2 How to Gamble With Little Regret

(a) If the deterministic strategy chooses expert $i(t)$ on day t, the adversary can assign 1 to the cost of expert $i(t)$'s advice, and 0 to the cost of other experts'. In other words, $c_{i(t)}^t = 1$ and $c_j^t = 0$ for $j \neq i(t)$. Then, the regret

$$R^* = \frac{1}{T}\left(T - \min_{1\leqslant j\leqslant n} \sum_{t=1}^{T} \delta_{j,i(t)}\right), \tag{1}$$

where δ_{kl} is Kronecker Delta.

Now we show that R^* is the maximum possible regret that the adversary can guarantee. Consider $c_{i(t)}^t + \delta_{j,i(t)} - c_j^t$ for any j. If $j = i(t)$, then that expression equals 1. Otherwise, it equals $c_{i(t)}^t - c_j^t$. For both cases, $c_{i(t)}^t + \delta_{j,i(t)} - c_j^t \leqslant 1$, which implies that

$$\sum_{t=1}^{T}[c_{i(t)}^t + \delta_{j,i(t)} - c_j^t] \leqslant T,\ \forall j = 1, 2, \cdots, n. \tag{2}$$

So,

$$\sum_{t=1}^{T}[c_{i(t)}^t - c_j^t] + \min_{1\leqslant k\leqslant n} \sum_{t=1}^{T} \delta_{k,i(t)} \leqslant T,\ \forall j = 1, 2, \cdots, n. \tag{3}$$

That is,

$$\sum_{t=1}^{T} c_{i(t)}^t - \min_{1\leqslant j\leqslant n} \sum_{t=1}^{T} c_j^t \leqslant T - \min_{1\leqslant k\leqslant n} \sum_{t=1}^{T} \delta_{k,i(t)} = T \cdot R^*. \tag{4}$$

As a result, R^* is the maximum possible regret given the deterministic strategy $i(t)$ ahead.

(b) Without loss of generality, suppose $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$. Now,

$$\mathbb{E}(R) = \frac{1}{T}\left(\sum_{t=1}^{T}\sum_{i=1}^{n} p_i \cdot c_i^t - \min_{1\leqslant j\leqslant n} \sum_{t=1}^{T} c_j^t\right). \tag{5}$$

The adversary can assign 0 to the cost of expert n's advice, and 1 to the cost of other experts' every day. If so, the expected regret is $(1 - p_n)$.

We show that $(1-p_n)$ is the maximum. Suppose $j^* \in \operatorname*{argmin}_{1\leqslant j\leqslant n} \sum_{t=1}^{T} c_j^t$. Then,

$$
\begin{aligned}
\mathbb{E}(R) &= \frac{1}{T}\left(\sum_{t=1}^{T}\sum_{i\neq j^*} p_i\cdot c_i^t - \sum_{t=1}^{T}(1-p_{j^*})c_{j^*}^t\right) \\
&\geqslant \frac{1}{T}\left(\sum_{t=1}^{T}\sum_{i\neq j^*} p_i\cdot 1 - \sum_{t=1}^{T}(1-p_{j^*})\cdot 0\right) \\
&\geqslant \frac{1}{T}\sum_{t=1}^{T}\sum_{i\neq j^*} p_i \\
&\geqslant 1-p_n.
\end{aligned}
\tag{6}
$$

To minimize this regret, the player needs to set p_n as high as possible. So, $p_1 = p_2 = \cdots = p_n = 1/n$.

3 Variants on the Experts Problem

(a) Let d_t denote the number of experts who have been correct on day t. We assume $d_0 = n$. When we make predictions on day t for day $t+1$, we guess the prediction of the majority of these d_t experts. First note that $d_{t+1} \leqslant d_t$. Besides, if we make a mistake on day t, $d_{t+1} \leqslant d_t/2$, since at least half of those d_t experts predict falsely. When d_t decreased to 1, we will not make mistakes anymore. So,

$$d_0 \geqslant 2^{\#\text{ of mistakes we make}} \tag{7}$$

which implies that we make at most $\log_2 n$ mistakes using this strategy.

(b) Similar to (a),

$$d_0 \geqslant k \cdot 2^{\#\text{ of mistakes we make}}. \tag{8}$$

So, we make at most $\log_2(n/k)$ mistakes using previous algorithm.

(c) We use x_i denote the first day when all experts have made at least i mistakes, $i = 1, 2, \cdots$. That is, on day $x_i - 1$, some expert has made only $(i-1)$ mistakes, but on day x_i, all experts have made at least i mistakes. If there is an expert who never makes i or even more mistakes, we set $x_i = +\infty$.

We prove by induction on l that we make at most $l \log_2 n$ mistakes before day x_l if x_l is finite. When $l = 1$, note that before day x_1, we only depend on predictions made by experts who have been correct. From (a), we know that we made at most $\log_2 n$ mistakes before day x_1.

Suppose the proposition holds for $l-1$, i.e. we make at most $(l-1)\log_2 n$ mistakes before day x_{l-1}. Between day x_{l-1} and day x_l, we only depend on predictions made by experts who have made exactly $(l-1)$ mistakes. On day x_{l-1}, there are at most n of them. If we make a mistake on some day in that period, the number of experts who have made exactly $(l-1)$ mistakes reduces by at least half. So, we make at most $\log_2 n$ mistakes before the number goes to zero, i.e., we are on day x_l. We make at most $\log_2 n$ mistakes between day x_{l-1} and day x_l. By the reduction hypothesis, we make at most $l \log_2 n$ mistakes before day x_l.

Now, come back to original problem. It is guaranteed that there is one expert who makes at most k mistakes. By the statement we proved above, we have made at most $k \log_2 n$ mistakes before day x_k. After day x_k, we will only depend on predictions made by experts who have made exactly k mistakes. Similar to previous proof, after day x_k we will make at most $\log_2 n$ mistakes, so we make at most $(k+1)\log_2 n$ mistakes in total.

It is easy for readers to construct a scenario where we make $(k+1)\log_2 n$ mistakes totally.

4 Weighted Rock-Paper-Scissors

(a) Suppose the maximizer is the row player. The payoff matrix for this game is:

		Column		
		r	p	s
Row	r	0	-2	1
	p	2	0	-4
	s	-1	4	0

(b) Suppose I play r, p and s with probability x_1, x_2 and x_3 respectively. My constrains are:

$$\begin{aligned} -2x_2 + x_3 + z &\leqslant 0 \\ 2x_1 - 4x_3 + z &\leqslant 0 \\ -x_1 + 4x_2 + z &\leqslant 0 \\ x_1 + x_2 + x_3 &= 1 \\ x_1, x_2, x_3 &\geqslant 0 \end{aligned} \tag{9}$$

My objective is

$$\max\ z. \tag{10}$$

After solving the problem, we get the maximum $z = 0$ when $x_1 = \frac{4}{7}, x_2 = \frac{1}{7}, x_3 = \frac{2}{7}$.

(c) My constrains are:

$$\begin{aligned} 10x_1 - 4x_2 - 6x_3 + z &\leqslant 0 \\ -3x_1 + x_2 + 9x_3 + z &\leqslant 0 \\ -3x_1 + 3x_2 - 2x_3 + z &\leqslant 0 \\ x_1 + x_2 + x_3 &= 1 \\ x_1, x_2, x_3 &\geqslant 0 \end{aligned} \tag{11}$$

My objective is

$$\max\ z. \tag{12}$$

The optimal strategy is to play rock with probability $x_1 = \frac{85}{254} = 0.3346457$, play paper with probability $x_2 = \frac{143}{254} = 0.5629921$ and play scissors with probability $x_3 = \frac{26}{254} = 0.1023622$. The expected payoff is $-\frac{122}{254} = -0.480315$.

(d) Suppose my friend plays r, p and s with probability y_1, y_2 and y_3 respectively. Their constrains are:

$$\begin{aligned} 10y_1 - 3y_2 - 3y_3 + z &\geqslant 0 \\ -4y_1 + y_2 + 3y_3 + z &\geqslant 0 \\ -6y_1 + 9y_2 - 2y_3 + z &\geqslant 0 \\ y_1 + y_2 + y_3 &= 1 \\ y_1, y_2, y_3 &\geqslant 0 \end{aligned} \tag{13}$$

Their objective is

$$\min\ z. \tag{14}$$

The optimal strategy is to play rock with probability $y_1 = \frac{68}{254} = 0.2677165$, play paper with probability $y_2 = \frac{82}{254} = 0.3228346$ and play scissors with probability $y_3 = \frac{104}{254} = 0.4094488$. The expected payoff is $-\frac{122}{254} = -0.480315$, which equals what we get in (c).

5 Domination

(a) The probability that the row player picks D is zero. To justify it, note that regardless of the column player's strategy, the row player can get better playoff by playing E than playing D.

(b) The probability that the column player picks A is zero. If she plays A with probability $p > 0$, she can further minimize the row player's payoff by playing B and C with additional probability $3p/4$ and $p/4$, because

$$\frac{3}{4}\begin{pmatrix}2\\2\\-2\end{pmatrix}+\frac{1}{4}\begin{pmatrix}-3\\-2\\2\end{pmatrix}=\begin{pmatrix}0.75\\1\\-1\end{pmatrix}\leqslant\begin{pmatrix}1\\3\\-1\end{pmatrix}. \tag{15}$$

Since the column player does not play A, the row player will never play D, since he may play E to get better payoff.

Another reason is that, given that the row player only plays E and F, the column player can get better playoff by playing B than playing A.

(c) Now, the row player only plays E and F, and the column player only plays B and C. It is easy to see that the optimal strategy is to play E and F with probability 0.5 and 0.5 for the row player, and to play B and C with probability 0.5 and 0.5 for the column player.

6 Follow the regularized leader

(a) **Follow the leader**

Suppose the payoff for picking strategy 1 is

$$A(t,1)=\begin{cases}0, & \text{if } t \text{ is odd;}\\ 1-\varepsilon^t, & \text{if } t \text{ is even.}\end{cases} \tag{16}$$

And the payoff for picking strategy 2 is

$$A(t,2)=\begin{cases}1-\varepsilon^t, & \text{if } t \text{ is odd;}\\ 0, & \text{if } t \text{ is even,}\end{cases} \tag{17}$$

where ε is a small positive number. On an even iteration $t = 2k$, the average payoff of strategy 1 on the previous iterations is

$$\frac{1}{2k-1}[(1-\varepsilon^2)+(1-\varepsilon^4)+\cdots+(1-\varepsilon^{2k-2})]=\frac{(k-1)-(\varepsilon^2+\varepsilon^4+\cdots+\varepsilon^{2k-2})}{2k-1}. \tag{18}$$

And the average payoff of strategy 2 on the previous iterations is

$$\frac{1}{2k-1}[(1-\varepsilon)+(1-\varepsilon^3)+\cdots+(1-\varepsilon^{2k-1})]=\frac{k-(\varepsilon^1+\varepsilon^3+\cdots+\varepsilon^{2k-1})}{2k-1}. \tag{19}$$

Since $\varepsilon \approx 0$, strategy 2 gives higher average payoff. I will pick strategy 2 on even iterations.

On an odd iteration $t = 2k+1 (k \geqslant 1)$, the average payoff of strategy 1 on the previous $2k$ iterations is

$$\frac{1}{2k}[(1-\varepsilon^2)+(1-\varepsilon^4)+\cdots+(1-\varepsilon^{2k})]=\frac{k-(\varepsilon^2+\varepsilon^4+\cdots+\varepsilon^{2k})}{2k}. \tag{20}$$

And the average payoff of strategy 2 on the previous $2k$ iterations is

$$\frac{1}{2k}[(1-\varepsilon)+(1-\varepsilon^3)+\cdots+(1-\varepsilon^{2k-1})]=\frac{k-(\varepsilon^1+\varepsilon^3+\cdots+\varepsilon^{2k-1})}{2k}. \tag{21}$$

Since $\varepsilon^1+\varepsilon^3+\cdots+\varepsilon^{2k-1} > \varepsilon^2+\varepsilon^4+\cdots+\varepsilon^{2k}$, strategy 1 gives higher average payoff. I will pick strategy 1 on odd iterations.

In conclusion, I pick strategy 1 on odd iterations, and strategy 2 on even iterations. Unfortunately, my payoff is 0. By eq. (20) and (21), however, sticking to either i = 1 or i = 2 would have given you a payoff of almost 50.

(b) **Follow the randomized leader**

To maximize

$$\sum_{i=1}^{n}\left(p_t(i)\cdot\sum_{\tau=1}^{t-1}A(\tau,i)\right)$$

at time t, one just sets $p_t(i) = 1$ for the strategy which maximize $\sum_{\tau=1}^{t-1} A(\tau,\cdot)$. This is the same as follow the strategy which gave the highest average payoff so far.

(c) **Follow the regularized leader** Suppose p_i and q_i are discrete probability distributions defined on the same sample space, $\mathcal{X} = 1, 2, \cdots, n$. By Jensen's inequality:

$$\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leqslant \ln \sum_{i=1}^{n} p_i \frac{q_i}{p_i} = \ln 1 = 0. \tag{22}$$

Let $y_1, y_2, \cdots, y_n$ be arbitrary positive real numbers. Let $q_i = y_i / \sum\limits_{j=1}^{n} y_j$, we have

$$\sum_{i=1}^{n} \left[p_i \ln \left(y_i / \sum_{j=1}^{n} y_j \right) - p_i \ln p_i \right] \leqslant 0.$$

So,

$$\sum_{i=1}^{n} (p_i \ln y_i - p_i \ln p_i) \leqslant \sum_{i=1}^{n} p_i \ln \sum_{j=1}^{n} y_j = \ln \sum_{i=1}^{n} y_i. \tag{23}$$

Let $y_i = \exp[\sum\limits_{\tau=1}^{t-1} A(\tau, i)/\eta]$, we have

$$\sum_{i=1}^{n} \left(p_i \frac{\sum\limits_{\tau=1}^{t-1} A(\tau, i)}{\eta} - p_i \ln p_i \right) \leqslant \ln \sum_{i=1}^{n} e^{\sum\limits_{\tau=1}^{t-1} A(\tau,i)/\eta},$$

i.e.

$$\sum_{i=1}^{n} \left(p_i \sum_{\tau=1}^{t-1} A(\tau, i) - \eta \cdot p_i \ln p_i \right) \leqslant \eta \ln \sum_{i=1}^{n} e^{\sum\limits_{\tau=1}^{t-1} A(\tau,i)/\eta}. \tag{24}$$

(d) Let $\varepsilon = 1 - e^{-1/\eta}$, so $1/(1-\varepsilon) = e^{1/\eta}$. For Multiplicative Weight Update,

$$w_i^{(t)} = \prod_{\tau=1}^{t-1} (1-\varepsilon)^{-A(\tau,i)} = (1-\varepsilon)^{-\sum\limits_{\tau=1}^{t-1} A(\tau,i)} = e^{\sum\limits_{\tau=1}^{t-1} A(\tau,i)/\eta} = y_i. \tag{25}$$

So, $p_i = w_i^{(t)} / \sum\limits_{j=1}^{n} y_j^{(t)} = y_i / \sum\limits_{j=1}^{n} y_j = q_i$. In this case, inequality (22) is tight, so inequality (24) is tight.