MVA 2020-21 P. Gaillard and R. Degenne

Sequential Learning

Final Examination

The duration of the exam is 2 hours. A single two-sided sheet of handwritten notes (with any content) is allowed. Answers can be written in French or English.

Part 1. Appetizers

1. Explain the difference between the pseudo regret and the regret of an online learning algorithm (in an adversarial setting).

2. (Doubling trick) Assume that an online learning algorithm $\mathcal{A}$ provides, for any known beforehand fixed horizon T, a regret bound $R_T(\mathcal{A}) \le CT^\alpha$ for some $C > 0$ and $\alpha > 0$. Explain how to convert it into an algorithm $\mathcal{A}_\infty$ which runs forever without knowing the horizon.

3. (UCB) Consider a stochastic bandit with K arms, distributions with support in $[0,1]$ and means $\mu(1), \dots, \mu(K)$. The UCB algorithm pulls arm $k_t = \arg\max \widehat{\mu}_t(k) + \sqrt{\frac{2\log t}{N_t(k)}}$, where $N_t(k) = \sum_{s=1}^{t-1} \mathbb{I}\{k_s = k\}$ is the number of pulls of arm k before t and $\widehat{\mu}_t(k)$ is an estimation of the mean of arm k. Suppose that for all $t \in \{1, \dots, T\}$, for all $k \in [K]$, $|\mu(k) - \widehat{\mu}_t(k)| \le \sqrt{\frac{2\log t}{N_t(k)}}$.

   (a) Show that
   $$N_t(k_t) \le \frac{8\log t}{\Delta_k^2}\,.$$

   (b) Prove an upper bound on the regret $R_T = T\max_{k\in[K]} \mu(k) - \sum_{t=1}^T \mu(k_t)$.

4. (FTL) Prove that there exists stochastic bandit problems with distributions with support in $[0,1]$ on which the Follow-The-Leader algorithm has linear expected regret $\mathbb{E}R_T = T\max_{k\in[K]} \mu(k) - \mathbb{E}\sum_{t=1}^T \mu(k_t)$ (where $\mu(k)$ is the mean reward of arm k).

Part 2. Successive rejects for best arm identification

We assume there are K unknown distributions $\nu(k)$ over $[0,1]$ with mean $\mu(k)$ for $k \in [K]$. For each arm k, $X_1(k), \dots, X_T(k)$ are i.i.d. random variables with distribution $\nu(k)$. When the player pulls arm k for the n^{th} time, the environment returns the reward $X_n(k)$ and the player observes that reward. We define $\widehat{X}_n(k) \stackrel{\text{def}}{=} (1/n)\sum_{m=1}^n X_m(k)$. We call k^* the optimal arm, i.e., $k^* \in \arg\max_{k\in[K]} \mu(k)$. We suppose that k^* is unique. We define $\overline{\log}(K) = \frac{1}{2} + \sum_{i=2}^K \frac{1}{i}$. We define $n_0 = 0$ and for $j \in \{1, \dots, K-1\}$, $n_j = \left\lceil \frac{1}{\overline{\log}(K)} \frac{T-K}{K+1-j} \right\rceil$.

Initialization: the set of active arms is $A_1 = \{1, \dots, K\}$.

For phases $j = 1, \dots, K-1$
- for all $k \in A_j$, pull arm k for $n_j - n_{j-1}$ times,
- compute $k_j \in \arg\min_{k\in A_j} \widehat{X}_{n_j}(k)$,
- deactivate arm k_j: the set of active arms becomes $A_{k+1} = A_k \setminus \{k_j\}$.

Recommend $\widehat{k}$, the only element of A_K.

Algorithm 1: Successive rejects algorithm

5. Algorithm 1 is designed for best arm identification with budget T. Prove that the total number of pulls of algorithm 1 is not larger than T.

6. We consider the 2-armed stochastic bandit framework, i.e. $K = 2$. Suppose without loss of generality that $k^* = 1$.

   (a) Prove that for $\alpha \geq 0$,

   $$\mathbb{P}\Big(\big|\widehat{X}_{n_1}(1) - \widehat{X}_{n_1}(2) - \mu(1) + \mu(2)\big| > \alpha\Big) \leq 2e^{-n_1\alpha^2/2}.$$

   (b) When it stops, the algorithm recommends $\widehat{k}$, the only arm in A_2. Let $\Delta = \mu(1) - \mu(2) > 0$. Prove that

   $$\mathbb{P}\Big(\widehat{k} \neq k^*\Big) \leq 2e^{-n_1\Delta^2/2}.$$

7. We now consider the general case $K \geq 2$. We suppose without loss of generality that $\mu(1) > \mu(2) \geq \ldots \geq \mu(K)$ and we define $\Delta_k = \mu(1) - \mu(k)$ for $k \in \{2, \ldots, K\}$.
   (a) What is the number of pulls of arm k at the end of phase j, if $k \in A_j$?
   (b) Prove that if $1 \in A_j$ and $1 \notin A_{j+1}$, then $\widehat{X}_{n_j}(1) \leq \max_{k \in \{K+1-j, \ldots, K\}} \widehat{X}_{n_j}(k)$. Note that even if the algorithm pulls arm k less than n_j times, $\widehat{X}_{n_j}(k)$ is still defined.
   (c) Deduce from the previous question that the probability that algorithm 1 recommends $\widehat{k} \neq 1$ is

$$\mathbb{P}\Big(\widehat{k} \neq 1\Big) \leq 2 \sum_{j=1}^{K-1} \sum_{k=K+1-j}^{K} e^{-n_j \Delta_k^2/2} \leq 2 \sum_{j=1}^{K-1} j e^{-n_j \Delta_{K+1-j}^2/2} .$$

   (d) Let $H_2 = \max_{k \geq 2} \frac{k}{\Delta_k^2}$. Prove that

$$\mathbb{P}\Big(\widehat{k} \neq 1\Big) \leq 2 \frac{K(K-1)}{2} e^{-\frac{T-K}{\log(K) H_2}} .$$

Part 3. Regularized follow the leader (RFTL)

Let $\Theta \subseteq \mathbb{R}^d$ be a compact convex decision space and $\eta > 0$. We consider the following setting. At each $t \geq 1$, the learner chooses $\theta_t \in \Theta$, then the environment chooses a convex loss $\ell_t : \Theta \to \mathbb{R}$ and reveals it to the learner. The goal of the learner is to minimize his regret

$$\mathrm{Regret}_T(\theta) = \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(\theta), \qquad \forall \theta \in \Theta .$$

We consider the RFTL algorithm defined in Algorithm 2, which depends on a strongly convex, smooth, and twice differentiable regularization function $R : \theta \to \mathbb{R}$.

8. Recall the gradient trick and provide a corresponding upper-bound on the regret.
9. (a) Show by induction that $\frac{R(\theta)}{\eta} + \sum_{t=1}^{T} g_t^\top \theta \geq \frac{R(\theta_1)}{\eta} + \sum_{t=1}^{T} g_t^\top \theta_{t+1}$ for any $\theta \in \Theta$.

Input: $\eta > 0$, regularization function $R > 0$, and a compact and convex set $\Theta \subset \mathbb{R}^d$
Let $\theta_1 = \arg\min_{\theta\in\Theta} \big\{R(\theta)\big\}$
For $t = 1$ to T do
- Play θ_t and observe $g_t = \nabla \ell_t(\theta_t)$
- Update

$$\theta_{t+1} = \operatorname*{arg\,min}_{\theta\in\Theta} \left\{ \eta \sum_{s=1}^{t} g_s^\top \theta + R(\theta) \right\}$$

end for

Algorithm 2: Regularized Follow the Leader

(b) Show that for any $\theta \in \Theta$

$$\mathrm{Regret}_T(\theta) \le \sum_{t=1}^{T} g_t^\top \big(\theta_t - \theta_{t+1}\big) + \frac{R(\theta) - R(\theta_1)}{\eta} \,.$$

We recall that the Bregman divergence $B_R(\theta||\theta')$ with respect to the function R is defined as

$$B_R(\theta||\theta') = R(\theta) - R(\theta') - \nabla R(\theta')^\top (\theta - \theta') \,.$$

We admit that for each $t \ge 1$, there exists a local norm $\|\cdot\|_t$ such that $B_R(\theta_t||\theta_{t+1}) = \frac{1}{2}\|\theta_t - \theta_{t+1}\|_t^2$, and we denote by $\|\cdot\|_t^*$ its dual norm that satifies the generalized Cauchy-Schwarz inequality $x^\top y \le \|x\|_t^* \, \|y\|_t$ for all $x, y \in \mathbb{R}^d$.

10. (a) Compute $B_R(\theta||\theta')$, $\|\cdot\|_t$ and $\|\cdot\|_t^*$ for $R(\theta) = \frac{1}{2}\|\theta - \theta_0\|^2$.

(b) Show that $\phi_t(\theta_t) \ge \phi_t(\theta_{t+1}) + B_R(\theta_t||\theta_{t+1})$ where $\phi_t(\theta) = \eta \sum_{s=1}^t g_s^\top \theta + R(\theta)$.

(c) Deduce that $B_R(\theta_t||\theta_{t+1}) \le \eta g_t^\top (\theta_t - \theta_{t+1})$.

(d) Show that $g_t^\top (\theta_t - \theta_{t+1}) \le 2\eta \|g_t\|_t^{*2}$.

(e) Let $G_R, D_R > 0$ such that for all $t \ge 1$, $\|g_t\|_t^* \le G_R$ and $\max_{\theta,\theta'\in\Theta}\{R(\theta) - R(\theta')\} \le D_R^2$. Show that for any $\theta \in \Theta$

$$\mathrm{Regret}_T(\theta) \le 2 D_R G_R \sqrt{2T}$$

for a well-chosen parameter $\eta > 0$ that needs to be explicited.