MVA 2019-20 Pierre Gaillard

Predictions of Individual Sequences

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.

The notation $\lesssim$ can be used to say "less or equal than, up to some universal (additive or multiplicative) constant".

Part 1. Appetizers

Let $\mathcal{A}$ be any online learning algorithm.

1. Explain the difference between the pseudo regret and the regret of $\mathcal{A}$.

2. Assume that the losses are in $[-1,1]$ and that for all $\delta > 0$ the regret of $\mathcal{A}$ is upper-bounded as $R_T(\mathcal{A}) \leq C\log(1/\delta)$, for some $C > 0$ with probability at least $1-\delta$. Prove the upper-bound on the expected regret $\mathbb{E}[R_T(\mathcal{A})] \leq C$.

3. (Doubling trick) Assume that $\mathcal{A}$ provides, for any known beforehand fixed horizon T, a regret bound $R_T(\mathcal{A}) \leq CT^\alpha$ for some $C > 0$ and $\alpha > 0$.
   (a) Explain how to convert it into an algorithm $\mathcal{A}_\infty$ which runs forever without knowing the horizon.
   (b) Prove the corresponding regret bound.
   (c) What regret bound do you get if $R_T(\mathcal{A}) = C\log T$?

4. (Linear regret) Assume that the algorithm $\mathcal{A}$ is determinist.
   (a) If $\mathcal{A}$ is asked to provide at each round $t \geq 1$, an action $\theta_t \in \{1,\dots,K\}$. Prove that there exists a sequence of loss functions $f_1,\dots,f_T \in [0,1]^K$ such that

$$R_T(\mathcal{A}) = \sum_{t=1}^T f_t(\theta_t) - \min_{k\in[K]} \sum_{t=1}^T f_t(k) \geq \Big(1-\frac{1}{K}\Big)T\,.$$

(b) If $\mathcal{A}$ is asked to provide at each round $t \geq 1$, an action $\theta_t \in \Delta_K = \{\theta \in [0,1]^K : \|\theta\|_1 = 1\}$. Prove that there exists a sequence of linear loss functions $g_1, \dots, g_T \in [0,1]^K$ such that

$$\sum_{t=1}^T \theta_t^\top g_t - \sum_{t=1}^T \min_{\theta \in [K]} \theta^\top g_t \geq \Big(1 - \frac{1}{K}\Big) T \,.$$

5. Show why Online Mirrored Descent with negative entropy $R(x) = \sum_{i=1}^K x(i) \log x(i)$ as regularization is equivalent to Exponential Gradient (Exponentially Weighted Average (EWA) forecaster with gradient trick).

Part 2. Successive elimination for stochastic bandits

For $t = 1, \dots, T$
- First, pull the two arms alternatively until there exists $k, k' \in \{1,2\}$ such that
  $$UCB_t(k) < LCB_t(k')$$
- Then, abandon arm k and play arm k' for the remaining rounds.

(a) $K = 2$

Initialization: all arms in $[K]$ are active.

For $t = 1, \dots, T$
- pull all active arms alternatively;
- if there exists k and $k' \in [K]$ such that
  $$UCB_t(k) < LCB_t(k')$$
  deactivate arm k.

(b) $K \geq 2$

Algorithm 1: Successive elimination algorithm

Setting and notations. We assume there are K unknown distributions $\nu(k)$ over $[0,1]$ with mean $\mu(k)$ for $k \in [K]$. At each time $t \geq 1$, the player chooses some action $k_t \in [K]$, the environment draws the reward $X_t(k_t) \sim \nu(k_t)$ and the player observes $X_t(k_t)$. We define $N_t(k) \stackrel{\text{def}}{=} \sum_{s=1}^t \mathbb{1}_{\{k_t = k\}}$ the number of times arm k was played before t and $\widehat{\mu}_t(k) \stackrel{\text{def}}{=} (1/N_t(k)) \sum_{s=1}^t X_t(k) \mathbb{1}_{\{k = k_t\}}$ the empirical mean of arm k at time t. We call k^* the optimal arm, i.e., $k^* \in \arg\max_{k \in [K]} \mu(k)$.

6. We consider the 2-armed stochastic bandit framework, i.e. $K = 2$ with Alg. 1(a).

(a) Prove that for $r_t(k) = \sqrt{2(\log T)/N_t(k)}$, we have

$$\mathbb{P}\Big(\forall t \in [T],\ \forall k \in [K], \quad \big|\widehat{\mu}_t(k) - \mu(k)\big| \le r_t(k)\Big) \le 1 - \frac{2}{T^2}\,.$$

(b) We define $UCB_t(k) \stackrel{\text{def}}{=} \widehat{\mu}_t(k) + r_t(k)$ and $LCB_t(k) \stackrel{\text{def}}{=} \widehat{\mu}_t(k) - r_t(k)$. Prove that

$$\bar{R}_T = \mathbb{E}\left[\max_{k\in[K]} \sum_{t=1}^{T} \mu(k) - \sum_{t=1}^{T} X_t(k_t)\right] \lesssim \sqrt{T \log T}\,.$$

7. We now consider the general case $K \ge 2$ (Alg. 1(b)). To ease the analysis, in the sequel, we consider the clean (high probability) event such that for all k, t, $\mu(k) \in [LCB_t(k), UCB_t(k)]$.

(a) Prove that for all suboptimal arms $k \in [K]$

$$\Delta(k) \stackrel{\text{def}}{=} \mu(k^*) - \mu(k) \lesssim r_T(k) \lesssim \sqrt{\frac{\log T}{N_T(k)}}\,.$$

In other words, a bad arm cannot be played too many times.

(b) Prove that therefore the pseudo regret satisfies $\bar{R}_T \lesssim \log T \sum_{k=1}^{K} \sqrt{N_T(k)}\,$.

(c) Conclude that $\bar{R}_T \lesssim \sqrt{KT\log T}$.

(d) Using the result of question 7a), show that we also have

$$\bar{R}_T \lesssim \sum_{k:\Delta(k)>0} \frac{\log T}{\Delta(k)}\,.$$

Part 3. Continuous Exponential Weights

Setting and notation. Let $\Theta \subseteq \mathbb{R}^d$ be a compact convex decision space and $\eta > 0$. A function $f : \Theta \mapsto \mathbb{R}$ is said η-exp-concave if $\theta \mapsto e^{-\eta f(\theta)}$ is concave over Θ. We consider the following setting of online prediction. At each round $t \ge 1$, the learner chooses $\theta_t \in \mathcal{X}$, then the environment chooses a continuous η-exp-concave loss $f_t : \Theta \to [0, 1]$ and reveals it to the learner.

8. Determine the set of η (if any) for which the following functions are η-exp-concave.

(a) the squared loss $f(\theta) = \|\theta - \theta^*\|_2^2$ for $\|\theta^*\|_2 \le B$.

(b) the ℓ_1-norm $f(\theta) = \|\theta - \theta^*\|_1$ for $\|\theta^*\|_1 \le B$.

9. We consider the continuous exponentially weighted average forecaster (EWA) that predicts

$$\theta_t = \frac{\int_{\theta\in\Theta} \theta w_t(\theta) d\mu(\theta)}{\int_{\theta\in\Theta} w_t(\theta) d\mu(\theta)}, \qquad \text{where} \qquad w_t(\theta) = \exp\Big(-\eta \sum_{s=1}^{t-1} f_s(\theta)\Big),$$

and where $d\mu$ is the uniform measure on Θ.

(a) Show that

$$W_T \stackrel{\text{def}}{=} \int_\Theta w_T(\theta) d\mu(\theta) \le \exp\left(-\eta \sum_{t=1}^T f_t(\theta_t)\right).$$

(b) Let $\varepsilon \in (0,1)$ and $\theta^* \in \arg\min_{\theta\in\Theta} \sum_{t=1}^T f_t(\theta)$. Define $\Theta_\varepsilon \stackrel{\text{def}}{=} \big\{(1-\varepsilon)\theta^* + \varepsilon\theta,\ \theta \in \Theta\big\}$. Show that for all $t \ge 1$ and all $\theta \in \Theta_\varepsilon$, we have

$$\exp\big(-\eta f_t(\theta)\big) \ge (1-\varepsilon)\exp\big(-\eta f_t(\theta^*)\big).$$

(c) Using that $\mu(\Theta_\varepsilon) \ge \varepsilon^d \mu(\Theta)$ (no proof needed), show that

$$W_T \ge \mu(\Theta)\varepsilon^d (1-\varepsilon)^T \exp\left(-\eta \sum_{t=1}^T f_t\big(\theta^*\big)\right).$$

(d) Conclude that the regret is upper-bounded as

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^T f_t(\theta_t) - \sum_{t=1}^T f_t(\theta^*) \lesssim \frac{d}{\eta}\log T.$$

(e) How would you implement this algorithm in practice and what is its complexity?