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.
This exam is made of 3 parts. The first part contains varied questions on the course. Parts 2 and 3 are exercices on adversarial and stochastic online learning respectively.

Part 1. Appetizers

1. Let $\Theta = \{1, \dots, K\}$ and $(\ell_t)_{1 \le t \le T}$ a sequence of adversarial losses from Θ to $[0,1]$. Consider the Follow-The-Leader strategy (FTL) that chooses $\widehat{\theta}_t \in \arg\min_{\theta \in \Theta} \big\{ \sum_{s=1}^{t-1} \ell_s(\theta) \big\}$. Provide a sequence of losses (ℓ_t) such that FTL incurs a regret larger than $(1 - 1/K)T - 1$.
2. What is the difference between a distribution-dependent and a distribution-free regret bound? What are the two corresponding bounds achieved by the Upper-Confidence-Bound algorithm?
3. Consider an online strategy $\mathcal{A}(g_1, \dots, g_{t-1}) = \widehat{\theta}_t$ that satisfies
$$\sup_{g_1,\dots,g_t \in B} \sup_{\theta \in B} \Big\{ \sum_{t=1}^T \langle g_t, \widehat{\theta}_t \rangle - \langle g_t, \theta \rangle \Big\} \le R_T, \qquad \text{where} \quad \mathcal{B} = \{x \in \mathbb{R}^d : \|x\|_2 \le G\}\,.$$
Explain how to convert $\mathcal{A}$ into to a strategy minimizing the regret with respect to convex L-Lipschitz losses $\ell_t : \mathcal{B} \to \mathbb{R}$.
4. In stochastic bandits, what are the drawbacks of the Explore-Then-Commit algorithm compared to UCB?

Part 2. Online Portfolio Selection

An investor starts with some initial wealth $S_0 = 1$, which she wants to invest over d assets. At each $t \ge 1$, she chooses a distribution $w_t \in \Delta_d$ over the assets. An adversary independently chooses market returns, i.e., a vector $x_t \in \mathbb{R}^d_{+,*}$ such that $x_t(i)$ is the price ratio of the i'th asset between rounds t and $t+1$. The ratio between the wealth of the investor at rounds t and $t+1$ is $\langle w_t, x_t \rangle$. Thus her total wealth after T rounds is $S_T = \prod_{t=1}^T \langle w_t, x_t \rangle$. Her goal is to maximize S_T which can be done by minimizing the regret

$$R_T := \sum_{t=1}^T \ell_t(w_t) - \sum_{t=1}^T \ell_t(w^*) = -\log \left(\frac{\prod_{t=1}^T \langle w_t, x_t \rangle}{\prod_{t=1}^T \langle w^*, x_t \rangle} \right),$$

where $\ell_t(w_t) = -\log(\langle w_t, x_t \rangle)$ and $w^* = \arg\max_{w \in \Delta_d} \prod_{t=1}^T \langle w, x_t \rangle$. For simplicity, we assume that $\|\nabla \ell_t(w)\| \le G$ for all w and t.

Input: $w_1 \in \Delta_d$, $\gamma > 0$, $A_0 = (4\gamma^2)^{-1} I_d$
For $t = 1$ to T do
– Play w_t and observe $g_t = \nabla \ell_t(w_t)$
– Update
$$w_{t+1} = \arg\min_{w \in \Delta_d} (\tilde{w}_{t+1} - w)^\top A_t (\tilde{w}_{t+1} - w) \tag{ONS}$$
where $\tilde{w}_{t+1} = w_t - \gamma^{-1} A_t^{-1} g_t$ and $A_t = A_{t-1} + g_t g_t^\top$
end for

Algorithm 1: Online Newton Step

5. (a) What would be the order (in G, d and T) of the regret of Online Gradient Descent?
   (b) Show that ℓ_t are exp-concave for all $t \geq 1$ and give the parameter.
   (c) How would you define a continuous version of the exponentially weighted average forecaster for this problem? What would be the order of its regret and the challenges to implement it?

6. The goal of the following questions is to show that Alg. 1 achieves a logarithmic regret bound.
   (a) Show that $A_t(\tilde{w}_{t+1} - w^*) = A_t(w_t - w^*) - \gamma^{-1} g_t$ and deduce
   $$(\tilde{w}_{t+1} - w^*)^\top A_t (\tilde{w}_{t+1} - w^*) = (w_t - w^*)^\top A_t (w_t - w^*) - \frac{2}{\gamma} g_t^\top (w_t - w^*) + \frac{1}{\gamma^2} g_t^\top A_t^{-1} g_t \,.$$
   (b) Show that it implies (and justify)
   $$g_t^\top (w_t - w^*) \leq \frac{1}{2\gamma} g_t^\top A_t^{-1} g_t + \frac{\gamma}{2}(w_t - w^*)^\top A_t (w_t - w^*) - \frac{\gamma}{2}(w_{t+1} - w^*)^\top A_t (w_{t+1} - w^*)$$
   (c) Because ℓ_t are expconcave, we **admit** that there exists $\gamma > 0$ such that for all $w, w' \in \Delta_d$ and $t \geq 1$
   $$\ell_t(w) \geq \ell_t(w') + \nabla \ell_t(w')^\top (w - w') + \frac{\gamma}{2}(w - w')^\top \nabla \ell_t(w') \nabla \ell_t(w')^\top (w - w') \,.$$
   Show that together with the previous question, it yields $R_T \leq \frac{1}{2\gamma} \sum_{t=1}^T g_t^\top A_t^{-1} g_t + \frac{1}{2\gamma}$.
   (d) Using that $\mathrm{Tr}(A^{-1}B) \leq \log\big(\det(A)/\det(A - B)\big)$ for any positive-semidefinite matrices $A \succ B \succcurlyeq 0$, prove that
   $$R_T \leq \frac{1}{2\gamma}\left(1 + \log \frac{\det(A_T)}{\det(A_0)}\right).$$
   (e) Provide a final regret bound in terms of T, G, γ and d only.

Part 3. Minimax regret in stochastic bandits

Consider a stochastic bandit problem in which $K > 1$ arms have Gaussian distributions with variance 1. The distribution of arm $k \in [K] = \{1, \dots, K\}$ is denoted by ν_k and has mean μ_k. We call such a bandit problem a "Gaussian bandit problem". At time $t \geq 1$, an algorithm picks an arm k_t, then observes a reward $X_t^{k_t}$. We define the regret at time T by

$$R_T = T \max_{j \in [K]} \mu_j - \sum_{t=1}^{T} \mu_{k_t} \,.$$

Algorithm 2 (UCB with known horizon T) is designed to minimize this regret. We denote by μ^* the maximal mean of an arm, $\mu^* = \max_{j \in [K]} \mu_j$, and denote the gap of arm $k \in [K]$ by $\Delta_k = \mu^* - \mu_k$.

For $t = 1$ to K do
- Pull arm $k_t = t$ and observe $X_t^{k_t} \sim \nu_{k_t}$
- Define $\widehat{\mu}_{K,k_t} = X_t^{k_t}$ and $N_{K,k_t} = 1$

end for
For $t = K+1$ to T do
- Compute $k_t = \arg\max_{k \in [K]} \widehat{\mu}_{t-1,k} + \sqrt{\frac{4 \log T}{N_{t-1,k}}}$
- Play arm k_t and observe $X_t^{k_t} \sim \nu_{k_t}$
- Define $N_{t,k_t} = N_{t-1,k_t} + 1$ and $N_{t,k} = N_{t-1,k}$ for $k \neq k_t$
- Define $\widehat{\mu}_{t,k_t} = \widehat{\mu}_{t-1,k_t} + \frac{1}{t}(X_t^{k_t} - \widehat{\mu}_{t-1,k_t})$ and $\widehat{\mu}_{t,k} = \widehat{\mu}_{t-1,k}$ for $k \neq k_t$

end for

Algorithm 2: Upper Confidence Bound (UCB) with known horizon T

7. The goal of this question is to prove a distribution-free regret bound for UCB (in the form shown in Algorithm 2).
   (a) Write the expected regret $\mathbb{E}[R_T]$ as an expression involving the gaps and the expected number of pulls $\mathbb{E}[N_{T,k}]$.
   (b) Show that for all $x \geq 0$, the expected regret of UCB is bounded from above by $Tx + \sum_{k:\Delta_k > x}(3\Delta_k + \frac{16 \log T}{x})$. You can use without proof that for all arms, $\mathbb{E}[N_{T,k}] \leq 3 + \frac{16 \log T}{\Delta_k^2}$.
   (c) Prove an upper bound on the expected regret of the form $\mathbb{E}R_T \leq Q(T, K) + 3\sum_{k=1}^{K} \Delta_k$, where $Q(T, K)$ is sub-linear in T and K and does not depend on the gaps.
8. Show that for any algorithm, either the expected regret verifies $\mathbb{E}R_T \geq \sum_{k=1}^{K} \Delta_k$ on all Gaussian bandit problems, or there exists one Gaussian bandit problem on which the algorithm has linear regret.
9. We will now prove lower bounds on the regret of any algorithm. Let $\Delta > 0$ and $\mu = (\Delta, 0, \dots, 0) \in \mathbb{R}^K$ be the vector of means of a Gaussian bandit problem, which we denote by ν. For $i \in \{2, \dots, K\}$, let $\mu^i = (\Delta, 0, \dots, 0, 2\Delta, 0, \dots, 0) \in \mathbb{R}^K$ (equal to μ except at coordinate i, where its value is 2Δ) be another mean vector. We call the corresponding Gaussian

bandit problem ν^i. We write $\mathbb{E}_\nu[...]$ for the expectation when the algorithm plays on problem ν, and $\mathbb{E}_{\nu^i}[...]$ the expectation on problem ν^i.

Let $j_{\min} = \arg\min_{k>1} \mathbb{E}_\nu[N_{T,k}]$ (any of them if the argmin is not unique).

(a) Prove that $\mathbb{E}_\nu[N_{T,j_{\min}}] \le \frac{T}{K-1}$.

(b) Prove that $\mathbb{E}_\nu[R_T] \ge \frac{T\Delta}{2}\mathbb{P}_\nu(N_{T,1} \le T/2)$ and that for any $i > 1$, $\mathbb{E}_{\nu^i}[R_T] \ge \frac{T\Delta}{2}\mathbb{P}_{\nu^i}(N_{T,1} > T/2)$.

(c) Let $H_T = (X_1^{k_1}, \dots, X_T^{k_T})$ be the history of observations up to time T. Let $\mathbb{P}_\nu^{H_T}$ be its distribution under problem ν and $\mathbb{P}_{\nu^i}^{H_T}$ be its distribution under problem ν^i. Give an expression of the Kullback-Leibler divergence $\mathrm{KL}(\mathbb{P}_\nu^{H_T}, \mathbb{P}_{\nu^i}^{H_T})$ which uses $\mathbb{E}_\nu[N_{T,k}]$ and $\mathrm{KL}(\nu_k, \nu_k^i)$ (Kullback-Leibler divergence between the Gaussian distributions of arm k under the two bandit problems) for all $k \in [K]$.

(d) Assume the following consequence of the Bretagnole-Huber inequality (and of the question above): for any event A and its complement A^c, $\mathbb{P}_\nu(A)+\mathbb{P}_{\nu^i}(A^c) \ge \exp\left(-\frac{1}{2}\mathbb{E}_\nu[N_{T,i}](\mu_i - \mu_i^i)^2\right)$. Prove that $\mathbb{E}_\nu[R_T] + \mathbb{E}_{\nu^{j_{\min}}}[R_T] \ge \frac{T\Delta}{2}\exp\left(-\frac{2T\Delta^2}{K-1}\right)$.

(e) Prove that there exists a constant C such that for $T \ge K$ and for any algorithm, there exists a Gaussian bandit problem ν' with mean vector $\mu' \in [0,1]^K$ such that

$$\mathbb{E}[R_T] \ge C\sqrt{(K-1)T}\,.$$