CPSC 532D, Fall 2024: Assignment 4
due Friday, December 6 at 11:59pm

You can do this with a partner if you'd like (there's a "find a group" post on Piazza). **Read the website section on academic integrity** here for what you're allowed to do and not do; in particular, cite your sources (including people you talked to!) and don't use ChatGPT/etc. If you're not sure if something is okay, ask.

Prepare your answers to these questions using LaTeX; hopefully you're reasonably familiar with it, but if not, try using Overleaf and looking around for tutorials online. Feel free to ask questions if you get stuck on things on Piazza (but remove any details about the actual answers to the questions... make a private post if that's tough). If you prefer, the `.tex` source for this file is available on the course website, and you can put your answers in `\begin{answer} My answer here... \end{answer}` environments to make them stand out; feel free to delete whatever boilerplate you want. Or answer in a fresh document.

Submit your answers as a single PDF on Gradescope: here's the link. Make sure to use the Gradescope group feature if you're working in a group. You'll be prompted to mark where each question is in your PDF; make sure you mark all relevant pages for each part (which saves us a surprising amount of grading time).

Please **put your name on the first page** as a backup, just in case. If something goes wrong, you can also email your assignment to me directly (`dsuth@cs.ubc.ca`).

1 Highwaymen, robbers, and so forth [55 + 5 challenge points]

We have an arm set $[K] = \{1, \ldots, K\}$. Rewards from pulling the jth arm are sampled from Bernoulli(μ_j), independent of everything else. Suppose there is a unique optimal arm j^*. Define $\Delta_j = \mu_{j^*} - \mu_j$ to be the gap in mean reward of arm j.

We will run the following game protocol for a sufficiently large number of rounds T (in particular, $T \gg K$).

In each round $t = 1, 2, \ldots, T$,

1. Nature decides upon a reward vector $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{K,t})$, by drawing each $X_{j,t} \sim$ Bernoulli(μ_j), independently of everything else.
2. Learner pulls an arm $J_t \in [K]$.
3. Learner obtains the reward $X_{J_t,t}$, and observes some information about X_t. (We'll vary the amount of information in the parts below.)

The regret of a given algorithm is given by

$$\mathcal{R}(T) := \max_{j \in [K]} \mathbb{E}\left[\sum_{t=1}^{T} X_{j,t}\right] - \mathbb{E}\left[\sum_{t=1}^{T} X_{J_t,t}\right].$$

We'll first consider the case where Learner gets to see the entirety of X_t after pulling arm J_t (full information).

[1.1] [30 points] Design a learning algorithm that achieves a $\mathcal{O}\left(\frac{\log K}{\Delta_{\min}}\right)$ problem-dependent regret bound, and $\mathcal{O}\left(\sqrt{T \log(K)}\right)$ worst-case regret bound. Here $\Delta_{\min} = \min_{j \in [K]: \Delta_j > 0} \Delta_j$ is the minimum mean reward gap. Note that the problem-dependent regret bound should not depend on T.

You also need to prove that your algorithm achieves these bounds. :)

Hint: Algorithm design: Since Learner can observe the reward vector X_t at the end of each round, exploration is not needed.

Hint: Problem-dependent regret analysis: It's easier to first assume all the sub-optimal arms have the same gap Δ, do that analysis, then use the doubling trick to estimate Δ_j and drop that assumption.

Hint: Worst-case regret analysis: you can set $\Delta_0 = \sqrt{\frac{\log K}{T}}$ to partition the arm set into two groups.

Answer: TODO

Now, let's consider a case with less than full information. Let's assume that K is an even number, and partition the arms into two disjoint groups: $C_1 = \{1, 2, \ldots, K/2\}$ and $C_2 = \{K/2 + 1, \ldots, K\}$.

In this version of the game, after pulling arm J_t, Learner gets to observe the values of $X_{j,t}$ for all j in the same group as J_t, including J_t itself. No information is provided about the rewards of arms in the other group.

[1.2] [25 points] Design a learning algorithm that achieves a $\sum_{g=1}^{2} \mathcal{O}\left(\frac{\log(T)}{\Delta_{g,\min}^2} \cdot \Delta_{g,\max}\right)$ problem-dependent regret bound, where $\Delta_{g,\min} = \min_{j \in C_g: \Delta_j > 0} \Delta_j$ is the minimum mean reward gap among all the sub-optimal arms in group C_g, and $\Delta_{g,\max} = \max_{j \in C_g} \Delta_j$ is the maximum mean reward gap among all the sub-optimal arms in group C_g.

Hint: Now, we do need to do some exploration, as Learner can only observe a group of arms in each round. However, within each group, pure exploitation is enough.

Answer: TODO

[1.3] [5 challenge points] Design a learning algorithm that achieves a $\sum_{g=1}^{2} \mathcal{O}\left(\frac{\log(T)}{\Delta_{g,\min}}\right)$ problem-dependent regret bound.

Hint: Use the doubling trick.

Answer: TODO

2 Piecewise-constant functions [35 points + 5 challenge points]

Let $a = (a_1, a_2, \ldots, a_k, 0, 0, \ldots)$ be an eventually-zero sequence with entries $a_i \in \{0, 1\}$. Then define a hypothesis $h_a : \mathbb{R}_{>0} \to \{0, 1\}$ by

$$h_a(x) = a_{\lceil x \rceil} = \begin{cases} a_1 & \text{if } 0 < x \leq 1 \\ a_2 & \text{if } 1 < x \leq 2 \\ & \vdots \end{cases}.$$

Consider the hypothesis class of all such functions: $\mathcal{H} = \{h_a : \forall i \in \mathbb{N}, a_i \in \{0, 1\} \text{ and } a \text{ is eventually zero}\}$. We'll use the 0-1 loss in this question.

[2.1] [5 points] Show $\operatorname{VCdim}(\mathcal{H}) = \infty$.

Answer: TODO

[2.2] [10 points] Give an example of a continuous distribution $\mathcal{D}_x$ on (a subset of) $\mathbb{R}_{>0}$ where, for some $m < \operatorname{VCdim}(\mathcal{H})$, samples $S_x \sim \mathcal{D}_x^m$ have probability zero of being shattered by $\mathcal{H}$. Thus prove that, for any $\mathcal{D}$ with this x marginal $\mathcal{D}_x$, ERM over $\mathcal{H}$ obtains error at most $\inf_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \varepsilon(m, \delta)$ with probability at least $1 - \delta$, where $\varepsilon(m, \delta)$ is some finite quantity such that $\lim_{m \to \infty} \varepsilon(m, \delta) = 0$ for each δ. By comparison, the VC bound would only show the approximation error is at most ∞.

Answer: TODO

[2.3] [10 points] Write $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots$, where each $\mathcal{H}_k$ has a finite VC dimension, and write down an explicit SRM algorithm that nonuniformly learns $\mathcal{H}$.

By "an explicit algorithm," I mean to expand out things like the uniform convergence bound for $\mathcal{H}_k$. It's okay to write something as an argmin over $\mathcal{H}$ like in equation (9.3) of the notes, if you say what k_h is for a given h and give the value of the corresponding Rademacher complexity. It's also okay to appeal to the SRM algorithm pseudocode from the notes, as long as you say what's in each $\mathcal{H}_k$, what the ε_k functions are, and how to compute the stopping condition.

Answer: TODO

[2.4] [3 challenge points] **Challenge question:** Suppose that instead of eventually-zero sequences, we allowed all possible sequences $a \in \{0, 1\}^{\mathbb{N}}$, e.g. the a that infinitely alternates between 0 and 1 is now an option. Prove that this bigger $\mathcal{H}'$ is *not* nonuniformly learnable. *This implies a sort of no-free-lunch theorem for nonuniform learnability.*

Hint: Try a diagonalization argument.

Answer: TODO

The following result will be useful momentarily:

Proposition 2.1. *Let $\mathcal{D}$ be any distribution over the positive integers $\mathbb{N}$, and $S \sim \mathcal{D}^m$. Define a random variable Q_S to be the number of unique samples seen out of m draws, $Q_S = \left|\{n : n \in S\}\right|$. Then $\mathbb{E}\, Q_S = o(m)$.*

(Recall little-o notation in this case is equivalent to saying $\lim_{m \to \infty} \frac{\mathbb{E}\, Q_S}{m} = 0$.)

[2.5] [5 points] Prove that, for *any* $\mathcal{D}_x$, $\mathbb{E}_{S_x \sim \mathcal{D}_x^m} \operatorname{Rad}(\mathcal{H}|_{S_x}) \to 0$ as $m \to \infty$.

Hint: You can use Proposition 2.1, if you reframe the problem slightly.

Answer: TODO

[2.6] [2 challenge points] **Challenge question:** Prove Proposition 2.1.

Answer: TODO

[2.7] [5 points] An absentminded professor once made the following argument on the *final exam* for a course:

If a hypothesis class has $\mathbb{E}_{S_x \sim \mathcal{D}_x^m} \operatorname{Rad}(\mathcal{H}|_{S_x}) \to 0$ for all $\mathcal{D}_x$, then for all realizable $\mathcal{D}$,

$$L_{\mathcal{D}}(\hat{h}_S) \le \mathop{\mathbb{E}}_{S_X \sim \mathcal{D}_x^m} \operatorname{Rad}(\mathcal{H}|_{S_x}) + \sqrt{\frac{1}{2m} \log \frac{1}{\delta}} \to 0.$$

Thus, by the "fundamental theorem of statistical learning," $\mathcal{H}$ must have finite VC dimension.

Clearly this argument is wrong, since it puts Questions [2.1] and [2.5] in contradiction. What was her mistake?

Answer: TODO