COMS 4995-1 S20 Homework 4 (due May 4, 2020)

Instructions

Submit your write-up on Gradescope as a neatly typeset PDF document by 11:00 PM of the due date. Please use TEX, LATEX, or a similar system.

On Gradescope, be sure to select the pages containing your answer for each problem. More details can be found on the Gradescope Student Workflow help page.

(If you don't select pages containing your answer to a problem, you'll receive a zero for that problem.)

Also, please make sure that your "Student ID #" on Gradescope is set to your UNI, using only lowercase letters and numbers (e.g., abc1234).

Finally, please make sure **your name and your UNI** appear prominently on the first page of your write-up.

Problem 1

In this problem, you will derive bounds on l^∞ covering numbers for linear functions in Δ^{d-1} (the probability simplex in $\mathbb{R}^d$) using the Hedge algorithm. (Recall that l^∞ covering numbers are always upper-bounds on l_n^2 covering numbers, so they can be plugged into the discretization lemma and Dudley's entropy integral.)

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ denote arbitrary vectors with $\|\boldsymbol{x}_i\|_\infty \leq 1$ for all $i = 1, \ldots, n$, and let

$$A := \left\{ (\boldsymbol{x}_1^\top \boldsymbol{p}, \ldots, \boldsymbol{x}_n^\top \boldsymbol{p}) : \boldsymbol{p} \in \Delta^{d-1} \right\}.$$

The goal is to prove a bound on $\mathcal{N}(\epsilon, A, l^\infty)$ for arbitrary $\epsilon > 0$. Recall that $\mathcal{N}(\epsilon, A, l^\infty) \leq N$ if there is a set $\tilde{A} \subset A$ of cardinality N that is an ϵ-cover of (A, l^∞), i.e., for all $\boldsymbol{a} \in A$, there exists $\tilde{\boldsymbol{a}} \in \tilde{A}$ such that $\|\boldsymbol{a} - \tilde{\boldsymbol{a}}\|_\infty \leq \epsilon$.

Recall that our analysis of Hedge from lecture assumes that $\boldsymbol{\ell}_t \in \mathbb{R}_+^d$, but it also works fine with arbitrary bounded loss vectors. The proof of the Hedge theorem can be modified to prove that if the loss vectors satisfy $\boldsymbol{\ell}_t \in [-1, +1]^d$, then using $\eta = \sqrt{\frac{2 \ln d}{T}}$ in Hedge guarantees, for all $\boldsymbol{p} \in \Delta^{d-1}$,

$$\sum_{t=1}^T \boldsymbol{\ell}_t^\top \boldsymbol{p}_t \leq \sum_{t=1}^T \boldsymbol{\ell}_t^\top \boldsymbol{p} + \sqrt{2T \ln d}.$$

Please check this yourself. You just need to prove that $\ln \frac{W_{t+1}}{W_t} \leq -\eta \boldsymbol{\ell}_t^\top \boldsymbol{p}_t + \frac{\eta^2}{2}$. And the trick to doing that is to think of it as the logarithm of a moment generating function...

(a) Consider the following procedure for an arbitrary input vector $\boldsymbol{q} \in \Delta^{d-1}$. Run the Hedge algorithm with $\eta = \sqrt{\frac{2 \ln d}{T}}$ for T rounds to produce $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T \in \Delta^{d-1}$, starting with $\boldsymbol{p}_1 = (1/d, \ldots, 1/d)$, on a sequence of loss vectors generated as follows. In round t, the loss vector $\boldsymbol{\ell}_t \in [-1, +1]^d$ is determined by first picking

$$i_t \in \arg\max_{i \in [n]} |\boldsymbol{x}_i^\top (\boldsymbol{p}_t - \boldsymbol{q})|,$$

and then defining

$$\boldsymbol{\ell}_t := \begin{cases} \boldsymbol{x}_{i_t} & \text{if } \boldsymbol{x}_{i_t}^\top (\boldsymbol{p}_t - \boldsymbol{q}) > 0; \\ -\boldsymbol{x}_{i_t} & \text{if } \boldsymbol{x}_{i_t}^\top (\boldsymbol{p}_t - \boldsymbol{q}) \leq 0. \end{cases}$$

After the T rounds, return

$$\bar{\boldsymbol{p}} := \frac{1}{T} \sum_{t=1}^T \boldsymbol{p}_t.$$

Prove that, for all $t = 1, \ldots, T$,

$$\max_{i \in [n]} |\boldsymbol{x}_i^\top (\boldsymbol{p}_t - \boldsymbol{q})| \leq \boldsymbol{\ell}_t^\top (\boldsymbol{p}_t - \boldsymbol{q}),$$

and also that

$$\max_{i \in [n]} |\boldsymbol{x}_i^\top (\bar{\boldsymbol{p}} - \boldsymbol{q})| \leq \sqrt{\frac{2 \ln d}{T}}.$$

(b) Imagine running the procedure from Part (b) for *all* possible $\boldsymbol{q} \in \Delta^{d-1}$, and let P be the set of vectors $\bar{\boldsymbol{p}}$ generated in that manner. Explain why $\tilde{A} := \{(\boldsymbol{x}_1^\top \boldsymbol{p}, \ldots, \boldsymbol{x}_n^\top \boldsymbol{p}) : \boldsymbol{p} \in P\}$ is an ϵ-cover for (A, l^∞) when $T = \lceil \frac{2 \ln d}{\epsilon^2} \rceil$, and prove a bound on the cardinality of $\tilde{A}$. Naturally, you should try to make your bound as small as possible.

Problem 2

Consider the following "online" version of the statistical learning setting. There is a hypothesis class $\mathcal{H}$ of hypotheses $h \colon \mathcal{X} \to \{\pm 1\}$ and an (unknown) probability distribution P over $\mathcal{X} \times \{\pm 1\}$. In round t, the learner must pick a hypothesis $h_t \in \mathcal{H}$, and then an independent example $(X_t, Y_t) \sim P$ is drawn and revealed to the learner. We say the learner makes a mistake in round t if $h_t(X_t) \neq Y_t$. Let M_T be the total number of mistakes made by the learner, and for any $h \in \mathcal{H}$, let $M_{T,h} = \sum_{t=1}^{T} \mathbb{1}_{\{h(X_t) \neq Y_t\}}$ be the total number of mistakes made by h. The regret after T rounds is

$$\mathrm{Reg}_T := M_T - \min_{h \in \mathcal{H}} M_{T,h}.$$

Assume you have access to an algorithm $\mathrm{ERM}_{\mathcal{H}}$ that, if provided any sequence of labeled examples from $\mathcal{X} \times \{\pm 1\}$, will return a hypothesis $h \in \mathcal{H}$ of minimum empirical risk over those examples. Also, for simplicity, assume that $\mathcal{H}$ has finite VC dimension $d \geq 1$.

(a) Consider the algorithm that, in round t, simply returns $h_t := \mathrm{ERM}_{\mathcal{H}}((X_1, Y_1), \ldots, (X_{t-1}, Y_{t-1}))$. Prove a bound of $O(\sqrt{dT})$ on the regret Reg_T after T rounds (that holds either in expectation, or with probability at least 0.99).

(b) Give an algorithm for the same setting that only calls $\mathrm{ERM}_{\mathcal{H}}$ at most $\log_2 T$ times over T rounds. Again, prove a bound of $O(\sqrt{dT})$ on the regret Reg_T after T rounds (that holds either in expectation, or with probability at least 0.99).

Problem 3

In this problem, you will finish the analysis of UCB by proving a "subtle" probability inequality.

The analysis of UCB needed a bound on the probability of an event of the following form:

$$\frac{1}{|S_t(a)|}\sum_{i\in S_t(a)} r_i(a) \geq \mu_a + \sqrt{\frac{\ln(1/\delta)}{|S_t(a)|}} \ \wedge\ |S_t(a)| \geq n_a.$$

Above, $r_i(a)$ is the random reward of action a in round i, which takes values in $[0,1]$, μ_a is the mean of $r_i(a)$, $S_t(a)$ is the set of rounds (up to t) in which the UCB algorithm chooses action a, and n_a is some positive number. Recall that, for a given action a, the random rewards are independent across rounds and have the same means. So, it seems plausible enough to bound the probability of this event as follows:

$$\begin{aligned}
&\Pr\left[\frac{1}{|S_t(a)|}\sum_{i\in S_t(a)} r_i(a) \geq \mu_a + \sqrt{\frac{\ln(1/\delta)}{|S_t(a)|}} \ \wedge\ |S_t(a)| \geq n_a\right] \\
&= \Pr\left[\frac{1}{|S_t(a)|}\sum_{i\in S_t(a)} (r_i(a)-\mu_a) \geq \sqrt{\frac{\ln(1/\delta)}{|S_t(a)|}} \ \middle|\ |S_t(a)| \geq n_a\right]\cdot \Pr\left[|S_t(a)|\geq n_a\right] \\
&\overset{???}{\leq} \delta\cdot\Pr\left[|S_t(a)|\geq n_a\right] \leq \delta.
\end{aligned}$$

However, the set $S_t(a)$ is, in fact, a random set. Because the action choices of the UCB algorithm depend on the random outcomes from previous rounds, the sum $\frac{1}{|S_t(a)|}\sum_{i\in S_t(a)} r_i(a)$ is not a sum of independent random variables. So the argument above is not valid.

Let us write the event with some more transparent notation. Let $X_i := \mathbb{1}_{\{a_i=a\}}$ be the indicator of the event that UCB chooses action a in round i, so $|S_t(a)| = \sum_{i=1}^t X_i$. Also let $Z_i := r_i(a) - \mu_a$. (Recall that $r_i(a)$ is a $[0,1]$-valued random variable.) Then, after some rearranging, the event we want to study is

$$\sum_{i=1}^t X_i Z_i \geq \sqrt{\sum_{i=1}^t X_i \ln(1/\delta)} \ \wedge\ \sum_{i=1}^t X_i \geq n_a.$$

The terms in the sum $\sum_{i=1}^t X_i Z_i$ are not independent: the X_i's depend on the (X_j, Z_j)'s for $j < i$.

What we are able to take advantage of is the following:

$$\mathbb{E}[X_i Z_i \mid X_{1:i}, Z_{1:i-1}] = 0.$$

This says that conditional on all of the outcomes from rounds prior to round i, as well as conditional on whether or not action a is taken in round i, the product $X_i Z_i$ has mean zero. This means that the $X_i Z_i$'s form a *martingale difference sequence*. This is great because there are many probability tools for dealing with martingale difference sequences (e.g., Freedman's inequality).

Below, we outline the steps to prove the desired probability inequality. Your task is to explain why each step holds. For example, for (d), you might write, "Step (d) follows from Step (c) by taking ln of both sides of the inequality inside the $\Pr[\cdot]$ and then dividing by λ." (Yes, many of the steps are rather simple.) Be as precise as possible in your explanations.

(a) For each $i = 1, \ldots, t$, and any $\lambda > 0$,

$$\mathbb{E}\left[\exp\left(\lambda X_i Z_i - \tfrac{\lambda^2}{8} X_i\right) \mid X_{1:i}, Z_{1:i-1}\right] \leq 1.$$

(b) For any $\lambda > 0$,

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{t} X_i Z_i - \frac{\lambda^2}{8} \sum_{i=1}^{t} X_i\right)\right] \leq 1.$$

(c) For any $\lambda > 0$ and $\delta \in (0, 1)$,

$$\Pr\left[\exp\left(\lambda \sum_{i=1}^{t} X_i Z_i - \frac{\lambda^2}{8} \sum_{i=1}^{t} X_i\right) \geq \frac{1}{\delta}\right] \leq \delta.$$

(d) For any $\lambda > 0$ and $\delta \in (0, 1)$,

$$\Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \frac{\lambda}{8} \sum_{i=1}^{t} X_i + \frac{1}{\lambda} \ln \frac{1}{\delta}\right] \leq \delta.$$

These first four steps (a)–(d) are essentially how Freedman proved his martingale tail inequalites.

The next several steps (e)–(j) use the following notation: $S_t := \sum_{i=1}^{t} X_i$ and $\lambda_j := \sqrt{\frac{8 \ln(1/\delta)}{\beta^{j+1} n_a}}$ for non-negative integers j. Here, $\beta > 1$ is a fixed constant. (The desired inequality uses $\beta = 2$.) Fix $\delta \in (0, 1)$. Then:

$$\begin{aligned}
&\Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \sqrt{\tfrac{\beta}{2} S_t \ln(1/\delta)} \;\wedge\; S_t \geq n_a\right] \\
&\overset{\text{(e)}}{=} \sum_{j=0}^{\log_\beta((t+1)/n_a)-1} \Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \sqrt{\tfrac{\beta}{2} S_t \ln(1/\delta)} \;\wedge\; \beta^j \leq S_t/n_a < \beta^{j+1}\right] \\
&\overset{\text{(f)}}{\leq} \sum_{j=0}^{\log_\beta((t+1)/n_a)-1} \Pr\left[\sum_{i=1}^{t} X_i Z_i \geq 2\sqrt{\tfrac{\beta^{j+1} n_a}{8} \ln(1/\delta)} \;\wedge\; \beta^j \leq S_t/n_a < \beta^{j+1}\right] \\
&\overset{\text{(g)}}{=} \sum_{j=0}^{\log_\beta((t+1)/n_a)-1} \Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \frac{\lambda_j}{8} \beta^{j+1} n_a + \frac{1}{\lambda_j} \ln(1/\delta) \;\wedge\; \beta^j \leq S_t/n_a < \beta^{j+1}\right] \\
&\overset{\text{(h)}}{\leq} \sum_{j=0}^{\log_\beta((t+1)/n_a)-1} \Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \frac{\lambda_j}{8} S_t + \frac{1}{\lambda_j} \ln(1/\delta) \;\wedge\; \beta^j \leq S_t/n_a < \beta^{j+1}\right] \\
&\overset{\text{(i)}}{\leq} \sum_{j=0}^{\log_\beta((t+1)/n_a)-1} \Pr\left[\sum_{i=1}^{t} X_i Z_i \geq \frac{\lambda_j}{8} S_t + \frac{1}{\lambda_j} \ln(1/\delta)\right] \\
&\overset{\text{(j)}}{\leq} \delta \log_\beta\left(\tfrac{t+1}{n_a}\right).
\end{aligned}$$

Note: This is definitely not the only way to analyze the UCB algorithm! For instance, the paper by Auer, Cesa-Bianchi, and Fischer (2002) uses a different argument.

Problem 4

In this problem, you will study the "Elimination" algorithm, which is a simple improvement over the "Explore Then Commit" algorithm in the Stochastic Bandits setting. (This problem is adapted from the forthcoming textbook, *Bandit Algorithms*, by Lattimore and Szepesvári.)

Elimination (adaptively) divides the rounds into "phases". The algorithm template is as follows.

- Initialize "active set" $A_1 := \{1, \ldots, N\}$.
- For phase $j = 1, 2, \ldots$:
  - Choose each action $a \in A_j$ exactly m_j times.
    - Note: This uses up $|A_j| m_j$ rounds...
  - Let $\mu_{j,a}$ be the empirical average of rewards for action a from this phase j.
  - Update active set:

$$A_{j+1} := \left\{ a \in A_j : \mu_{j,a} + 2^{-j} \geq \max_{a' \in A_j} \mu_{j,a'} \right\}.$$

Fix $a^* \in \arg\max_{a \in [N]} \mu_a$. Assume, for simplicity, that for each action a, the rewards across rounds are iid random variables taking values in $[0, 1]$.

(a) Show that there is an absolute constant $c_1 > 0$ such that for any $j \geq 1$,

$$\Pr(a^* \in A_j \wedge a^* \notin A_{j+1}) \leq N \exp\left(-c_1 m_j 2^{-2j}\right).$$

(b) Show that there is an absolute constant $c_2 > 0$ such that, for all $a \in [N]$ and $j \geq 1$ satisfying $\gamma_a \geq 2^{-j}$,

$$\Pr(a^* \in A_j \wedge a \in A_j \wedge a \in A_{j+1}) \leq \exp\left(-c_2 m_j (\gamma_a - 2^{-j})^2\right).$$

(c) Let $j_a := \min\{j \geq 1 : 2^{-j} \leq \gamma_a/2\}$ for each $a \in [N]$. Determine values for the m_j such that

$$\Pr(\exists j \geq 1 \text{ s.t. } a^* \notin A_j) \leq \frac{1}{T},$$

$$\Pr(a \in A_{j_a+1}) \leq \frac{1}{T} \quad \text{for all } a \in [N] \text{ such that } \gamma_a > 0,$$

(d) Prove that for choices of the m_j satisfying the properties from Part (c), there exists an absolute constant $c_3 > 0$ such that

$$\mathbb{E}\left[\sum_{t=1}^{T} \mu_{a^*} - \mu_{a_t}\right] \leq c_3 \sum_{a:\gamma_a > 0} \left(\gamma_a + \frac{\log T}{\gamma_a}\right).$$