Fall 2022 *Instructor: Jacob Abernethy, Bhevush Kumar, Guanghui Wang*

ECE 8803: Online Decision Making in Machine Learning

Homework 3

Released: Nov 3 *Due: Nov 16*

Problem 1 (Boosting via Game Playing) 20 points There is a popular framework for constructing a hypothesis via an *ensemble*, i.e. mixing together a collection of predictors, that is known as *boosting*. What boosting does is iteratively improve the ensemble by incorporating new predictors (aka "weak learners"). The problem, as it turns out, can be formulated as solving a zero-sum game, and the iterative process can be viewed as an interaction between a no-regret learning algorithm and an "oracle" for choosing predictors. Before we begin, let's lay out some terminology.

Let $\mathcal{X}$ be a data space (e.g. $\mathbb{R}^d$), and assume we have labels $\mathcal{Y} := \{-1, +1\}$. We are given access to n examples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathcal{X} \times \mathcal{X}$. We also have a set of *weak learners*, $\mathcal{H} = \{h : \mathcal{X} \to \{0, 1\}\}$, which make predictions. For the remainder of this problem, let's assume $\mathcal{H}$ is finite, $|\mathcal{H}| = m$. We typically refer to these predictors as "weak" because, in practice, they will have very low complexity. A common choice of weak learner class is the set of *decision stumps*, which simply predict +1/-1 based on whether a single feature is above/below a given threshold. We don't expect a single weak learner to perform well, but perhaps we can combine them into a "stronger" predictor? Here's a classic assumption we make about the weak learning class $\mathcal{H}$.

Weak learning assumption ($\gamma > 0$)*:* Assume, for every dist. $\mathbf{p} \in \Delta_n$, there exists $h \in \mathcal{H}$

$$\Pr_{i \sim \mathbf{p}}[h(x_i) \neq y_i] \leq \frac{1}{2} - \frac{\gamma}{2}$$

This assumption is "weak" in the sense that it only guarantees that some $h \in \mathcal{H}$ is going to perform slightly better than a 50-50 guess. Of course, what we would prefer is a strong predictor, which is 100% accurate. When might we achieve this with a mixture of weak learners? Let's define this as follows.

Strong learning assumption ($\gamma > 0$)*:* Assume there is some distribution $\mathbf{q} \in \Delta_m$ for which the $\mathbf{q}$-weighted mixture of predictors $\mathcal{H}$ always has a γ-margin for every example. That is, $\forall i \in [n]$

$$\text{The weighted forecast } \sum_{j=1}^{m} q_j h_j(x_i) \text{ is } \begin{cases} \geq \gamma & \text{when } y_i = +1 \\ \leq -\gamma & \text{when } y_i = -1 \end{cases}$$

Let's now try to reformulate these ideas in the form of a min-max problem.

(a) Let's try to rewire the weak learning assumption in matrix form. Construct an $n \times m$ matrix, with values in $\{-1, +1\}$, so that the weak learning assumption is equivalent to the statement

$$\min_{\mathbf{p} \in \Delta_n} \max_{j \in [m]} \mathbf{p}^\top M \mathbf{e}_j \geq \gamma,$$

where $\mathbf{e}_j$ is the jth basis vector (all zeros, except for a 1 in the jth coordinate).

(b) Argue why the above minmax statement is equivalent to the following

$$\min_{\mathbf{p}\in\Delta_n} \max_{\mathbf{q}\in\Delta_m} \mathbf{p}^\top M\mathbf{q} \geq \gamma.$$

(c) Similarly, argue why the strong learning assumption can be reformulated as

$$\max_{\mathbf{q}\in\Delta_m} \min_{\mathbf{p}\in\Delta_n} \mathbf{p}^\top M\mathbf{q} \geq \gamma.$$

(d) Using von Neumann's minimax theorem argue that, for a fixed $\gamma > 0$, the weak learning assumption implies the strong learning assumption.

(e) (BONUS - 15 points) Formulate an algorithm, using the exponential weights method, that finds a distribution over the weak learners $\mathbf{q}^* \in \Delta_m$ which satisfies the strong learning assumption for some $\gamma > 0$. Your algorithm should require roughly $T = O\left(\frac{\log n}{\gamma^2}\right)$ updates. Your technique should iteratively find a sequence of distributions $\mathbf{p}_t$, for $t = 1, \ldots, T$, using the exponential weights method, and then the "other player" should select a weak learner $h_{j_t} \in \mathcal{H}$ which satisfies $\mathbf{p}_t^\top M\mathbf{e}_{j_t} \geq \gamma$. The latter is always guaranteed via the weak learning assumption. Your analysis should follow the standard analysis of solving a zero-sum game using no-regret algorithms, as described in class, but you'll need to think carefully how to extract the final distribution $\mathbf{q}^*$. (*Note:* this algorithm is very close to the Adaboost algorithm! It's not exactly the same only because Adaboost has an additional adaptive parameter update which doesn't require it to know γ in advance.)

Problem 2 (Understanding the pseudo-regret of greedy, pure-exploration and explore-then-commit) 20 points In class, we introduced the two-armed Bernoulli bandit problem through the motivation of drug discovery. In particular, we considered a 2-armed Bernoulli bandit problem with reward parameters $p_1 = 0.2$ and $p_2 = 0.7$, corresponding to the mean efficacy of Drugs A and B respectively on a patient. We claimed that the greedy algorithm and "only-explore" algorithm were both highly suboptimal in terms of overall reward, but the explore-then-commit algorithm did better. In this problem, we will formally explore this through the metric of *pseudo-regret* (defined in Lectures 11 and 12).

All expectations and probabilities will be over the randomness in the reward sequence $\{G_{t,i}\}_{t\geq 1}$ for $i = 1, 2$. You will find it useful to review the notes in Lectures 11 and 12 to solve this problem.

(a) Consider the greedy algorithm with the default choice of picking arm 1 (drug A) on round 1, arm 2 (drug B) on round 2, and greedy thereafter. What is the probability that you will *always* pick arm 1 round 3 onwards?

(b) Use the above sub-part to *lower bound* the pseudo-regret of the greedy algorithm. What is the dependence of this lower bound on T?

Hint: Under the situation of part (a), what will the expected reward be from rounds 3 to T? How does this compare to the benchmark of $0.7(T-2)$?

(c) Now, consider the "pure-explore" algorithm, which picks arm 1 on odd rounds and arm 2 on even rounds. Calculate the exact pseudo-regret of this algorithm.

Hint: use the formula for pseudo-regret in terms of the expected number of times the suboptimal arm was sampled.

(d) Finally, consider the explore-then-commit (ETC) algorithm from Lecture 11, where the arms are sampled in a round-robin fashion for $T_0 < T$ rounds. Assume that T_0 is even. Use Hoeffding's bound to derive an *upper bound* on the probability that you will always pick the suboptimal arm 1 after T_0 rounds.

(e) Use part (e) to derive an upper bound for the pseudo-regret of ETC with T_0 exploration rounds. What is the value of T_0 that minimizes the upper bound? Use this choice to state an upper bound on pseudo-regret in terms of T, the total number of rounds.

(f) (BONUS – 10 points) Repeat all of the above parts for arbitrary values of p_1, p_2. Denote $\Delta = |p_2 - p_1|$ and express your eventual bounds on pseudo-regret as a function of Δ. Suppose that you did not know Δ beforehand. Then, what is the value of T_0 that minimizes pseudo-regret over all values of Δ?

Hint: use the fact that pseudo-regret is also upper-bounded by ΔT.

Problem 3 (The successive arm elimination algorithm) 20 points In this problem, we examine the psudo-regret of the successive arm elimination (SAE) algorithm for the 2-armed bandit problem with rewards bounded between $[0, 1]$. In class, we have already discussed the main idea of the this algorithm. Here, we will study its detailed proof. Without loss of generality, assume $\mu_1 < \mu_2$ (and $\mu_1, \mu_2 \in [0, 1]$).

This algorithm is very similar in some ways to the UCB algorithm, but a little different in its day-to-day behavior. Instead of constantly keeping all arms in play like UCB, it plays a current set of "active" arms in a round-robin fashion, and eliminates arms that seem to be performing suboptimally at the end of each round-robin turn. In more detail, the basic procedure of SAE at round t, when both arms are in play, is as follows:

- Define the upper confidence bound of arm $a \in \{1, 2\}$ as $\mathsf{UCB}(a, t) = \hat{\mu}_{a,t-1} + \sqrt{\frac{\log(1/\delta)}{2N_{t-1}(a)}}$, and the lower confidence bound $\mathsf{LCB}(a, t) = \hat{\mu}_{a,t-1} - \sqrt{\frac{\log(1/\delta)}{2N_{t-1}(a)}}$.
- If t is odd, play arm 1.
- If t is even:
  - Play arm 2.
  - If $\mathsf{UCB}(1, t+1) < \mathsf{LCB}(2, t+1)$, eliminate arm 1 and *play arm* 2 *on all rounds there-after.*
  - Else, if $\mathsf{UCB}(2, t+1) < \mathsf{LCB}(1, t+1)$, then eliminate arm 2 and *play arm* 1 *on all rounds there-after.*
  - If neither elimination criterion is met, then keep both arms in play and proceed to the next round.

(a) Set $\delta = 1/T^2$ (as we did for UCB), and define the "good event" $\mathcal{A}$ as $\mathcal{A} = \{\mu_i \in [\mathsf{LCB}(i, t), \mathsf{UCB}(i, t)]$ for all $i \in \{1, 2\}$ and all $t \in \{1, \ldots, T\}.\}$. Prove that $\mathcal{A}$ occurs with probability at least $1 - 4/T$.

Hint: Use Hoeffding's inequality together with the ideas in lecture note 14 (in particular, Section 14.1.1 is helpful).

(b) Let $\hat{t}$ denote the round on which we eliminate an arm. Conditioned on the "good event" $\mathcal{A}$, prove that the pseudo-regret incurred on rounds $t = \hat{t}+1, \ldots, T$ is equal to 0.

Hint: Use the elimination criterion to show that when the "good event" $\mathcal{A}$ holds, the eliminated arm must be the suboptimal arm 1. Drawing a picture of the confidence intervals $[\mathsf{LCB}, \mathsf{UCB}]$ for each arm might help.

(c) Next, let's consider the pseudo-regret for $t < \hat{t}$. According to the elimination rule, we know that round $\hat{t}-2$ is the last round that we do *not* drop any arm. Based on this fact, prove that $\Delta = |\mu_1 - \mu_2| \leq 4\sqrt{\frac{2\log(T)}{(\hat{t}-2)}}$ under the "good event" $\mathcal{A}$.

Hint: First, use the fact that the elimination criterion does not hold at round $\hat{t}-2$ to show that $\mathsf{UCB}(2, \hat{t}-2) - \mathsf{LCB}(1, \hat{t}-2) \leq 4\sqrt{\frac{2\log(T)}{(\hat{t}-2)}}$. Then, use the definition of the "good event" in part (a).

(d) Based on part (c), provide an upper-bound on the pseudo-regret of SAE from round 1 to round $\hat{t}$ conditioned on the "good event" $\mathcal{A}$ that depends only on T and Δ.

(e) Combine the above results and provide an upper-bound on the overall pseudo-regret for SAE from round 1 to round T (considering both the "good event" $\mathcal{A}$ and the "bad event" $\mathcal{A}^C$). Like part (d), your upper bound should depend only on T and Δ.