Online Learning with Sleeping Experts and Feedback Graphs

Corinna Cortes [1] **Giulia DeSalvo** [1] **Claudio Gentile** [1] **Mehryar Mohri** [1 2] **Scott Yang** [3]

Abstract

We consider the scenario of online learning with sleeping experts, where not all experts are available at each round, and analyze the general framework of learning with feedback graphs, where the loss observations associated with each expert are characterized by a graph. A critical assumption in this framework is that the loss observations and the set of sleeping experts at each round are independent. We first extend the classical sleeping experts algorithm of Kleinberg et al. (2008) to the feedback graphs scenario, and prove matching upper and lower bounds for the sleeping regret of the resulting algorithm under the independence assumption. Our main contribution is then to relax this assumption, present a more general notion of sleeping regret, and derive a general algorithm with strong theoretical guarantees. We apply this new framework to the important scenario of online learning with abstention, where a learner can elect to abstain from making a prediction at the price of a certain cost. We empirically validate our algorithm against multiple online abstention algorithms on several real-world datasets, showing substantial performance improvements.

1. Introduction

Sequential decision making under uncertainty is an important and widely studied area of machine learning. In the standard online learning framework (Cesa-Bianchi & Lugosi, 2006), at each round, the learner selects an action out of a finite set and incurs some loss associated with that action. The learner's goal is to minimize her regret over a finite number of rounds, that is the difference between her cumulative loss and that of the best static action in hindsight.

Online learning with feedback graphs is a general framework for online learning where the action losses that are observable to the learner are modelled by graphs. This framework was first introduced by Mannor & Shamir (2011) and later analyzed by several other authors (Caron et al. (2012); Buccapatnam et al. (2014); Wu et al. (2015)Alon et al. (2013); Kocák et al. (2014) Alon et al. (2015); Cohen et al. (2016); Kocák et al. (2016); Tossou et al. (2017); Liu et al. (2018); Yun et al. (2018)). Given a directed feedback graph, an edge from i to j indicates that the loss of j is observed if expert i is selected by the algorithm. Such partial observability setups cover a variety of applications (e.g. in web advertising, a user who clicks on an ad reveals information about related ads). The classical settings of full information (Littlestone & Warmuth, 1994) and bandit (Auer et al., 2002) online learning are special instances corresponding to a fully connected graph and a graph admitting only self-loops, respectively. In general, these graphs can be either fixed or time-varying, and they can also even be stochastic.

Distinct from the feedback graph framework, online learning has also been studied in a setting in which the actions available to the learner can change at different rounds. This scenario is called the sleeping experts setting. It was analyzed first by (Kleinberg et al., 2008) and subsequently by (Kanade et al., 2009; Kanade & Steinke, 2014). In this framework, at each round the environment determines a set of available actions either stochastically or adversarially. This model of online learning can arise, e.g., in routing network problems, where some routes may be unavailable due to either random router crashes (stochastic case) or an illicit agent (adversarial case).

Restricted feedback and restricted action sets are two closely related ideas, and many applications can actually be formalized as a combination of both sleeping experts and feedback graphs. For instance, consider again the scenario of web advertising described above, where a learner has to decide which ads to display. Some ads may not be available at each round, implying that the experts are sleeping, and at the same time, related ads may have similar rewards, so that the feedback between some of the ads should be shared. Similarly, in e-commerce from the seller's perspective, some items may be out of stock (sleeping experts), and the rewards for one item may be similar to the rewards of others (feedback graph). As a third example, sensor networks are

[1]Google Research, New York, NY; [2]Courant Institute of Mathematical Sciences, New York, NY; [3]D. E. Shaw & Co., New York, NY. Correspondence to: Giulia DeSalvo <giuliad@google.com>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

a common motivation for feedback graphs (e.g. (Mannor & Shamir, 2011)), where a centralized controller must activate a sensor to receive input from it and where the area covered by sensors tend to overlap. In this problem, some of the routes or links are down due to mechanical issues, which can be modelled using the sleeping experts framework.

It is natural in many problems for a restricted action set to also considerably impact a learner's feedback graph. At the same time, these two concepts are distinct, and neither one of the frameworks alone can capture the interplay between action set availability and feedback between actions. Moreover, these two concepts have heretofore been studied in isolation from one another. Online learning with feedback graphs has been studied only in the classical setting where the learner always has access to every action, and online learning with sleeping experts has only been studied in the full information and bandit settings.

In this work, we introduce and analyze an online learning setting admitting both feedback graphs and sleeping experts. We first consider a simpler scenario in which the losses and the awake sets are statistically independent, which, with either full information or bandit feedback, is the case of sleeping experts studied in prior work. We then move on to a more complicated setting in which the losses and the awake sets are dependent. As an application, we apply our ideas to the scenario of online learning with abstention recently introduced by Cortes et al. (2018). This is a setting, relevant in practice, where a learner can decide to abstain from making predictions at the price of a known cost and restricted feedback.

The paper is organized as follows. In Section 2, we define relevant notation and formally introduce the setting. In Section 3, we analyze the setting in which losses and awake sets are statistically independent. We extend the AUER algorithm, a standard sleeping experts algorithm, to incorporate the loss observations encoded by feedback graphs. We derive sleeping regret guarantees based on the expected loss gaps and feedback graph structure. In Section 4, we analyze the more complex scenario in which the loss observations and awake sets are not statistically independent. We first highlight deficiencies with the classical notion of sleeping regret, and we then propose a new and more informative quantity, which we call generalized sleeping regret. We then introduce a novel algorithm that uses both the awake sets and loss observations to estimate unbiased empirical losses. We present guarantees for our new notion of regret that is logarithmic in the number of rounds and that depends on conditional expected loss gaps. In Section 5, we show how our algorithms can be adapted to the abstention setting to yield more compelling theoretical guarantees than those in previous works. In Section 6, we corroborate our theoretical results for the abstention setting with extensive experiments against multiple online abstention algorithms on several real-world datasets, showing that substantial improvements are also achieved empirically.

2. Preliminaries

We denote by $\mathcal{X}$ the input space, by $\mathcal{Y}$ the output space, and by $\mathcal{D}$ a probability distribution over $\mathcal{X} \times \mathcal{Y}$. Let $\mathcal{E}$ denote the family of experts (or actions): $\mathcal{E} = \{\xi_j \colon j \in [K]\}$, where $[K] = \{1, \ldots, K\}$, and let $L \colon \mathcal{E} \times (\mathcal{X} \times \mathcal{Y}) \to [0, 1]$ be a loss function.

We consider the scenario of online learning with side-information modeled by *feedback graphs* introduced by Mannor & Shamir (2011). For any $t \in [T]$, a feedback graph $G^t = (V^t, E^t)$ is a directed graph over the set of experts with indices $i \in [K]$, which admits an edge from i to j if the loss of j is observed by the algorithm when it selects expert i at round t. Let N_i^t denote the out-neighborhood of i at time t, that is the set of vertices $j \in [K]$ for which G^t admits an edge from i to j. We will specify our assumptions behind how the graphs G^t are generated in future sections.

We consider a stochastic setting of online learning with feedback graphs, which admits the following learning protocol. At each round $t \in [T]$, a pair $(x_t, y_t) = z_t \in \mathcal{X} \times \mathcal{Y}$ is drawn i.i.d. from $\mathcal{D}$. The learner receives the input $x_t \in \mathcal{X}$ drawn i.i.d. according to the marginal distribution associated with $\mathcal{D}$, selects an index $I_t \in [K]$ corresponding to an expert $\xi_{I_t} \in \mathcal{E}$, incurs the loss $L(\xi_{I_t}, z_t)$, and observes the loss of every expert in the out-neighborhood of I_t, that is, $L(\xi_j, z_t)$, with $j \in N_{I_t}^t$. Note that the full information setting corresponds to the case where, for all t, G^t is the fully-connected graph, while the multi-armed bandit model matches the case where G^t only contains self-loops. In what follows, we assume that the loss of the expert selected is always observed. Thus, for all $t \in [T]$, G^t contains self-loops at all nodes: $i \in N_i^t$, for all $i \in [K]$.

We also adopt the *sleeping experts* framework introduced by Kleinberg et al. (2008). In this setting, at each round t, the environment also generates a set $A^t \subseteq [K]$ of available (or *awake*) experts. We denote by $\mathcal{A}$ the set of all awake sets that can be possibly generated, which is a subset of the power set of $[K]$.

Note that, while the sleeping expert and feedback graph frameworks are related, one does not subsume the other. Awake sets determine which experts can be chosen by the learner, while feedback graphs determine which losses can be observed by the learner. Thus, there is no distribution over awake sets that can mimic a feedback graph scenario, and there is no set of feedback graphs that can lead to a sleeping experts setting.

Since not all experts or actions are available at each round

in the sleeping experts setting, the best expert in hindsight, used in the standard notion of regret, is not a realistic benchmark in this setting. Instead, the notion of *sleeping regret* was introduced by Kleinberg et al. (2008).[1] This notion of regret considers the difference between the cumulative loss of the algorithm and that of the best ordering of the experts, where, at each round, the expert with the most favorable rank among those awake is selected.

In our setting, the sleeping regret of an algorithm ALG can be defined as follows:

$$\mathrm{r}_T^{\text{SLEEP}}(\text{ALG}) = \sum_{t=1}^{T} \mathbb{E}[L(\xi_{I_t}, z_t)] - \sum_{t=1}^{T} \mathbb{E}[L(\xi_{\sigma(A_t)}, z_t)],$$

where I_t is the index of the expert selected by ALG at round t, and $\sigma(A^t)$ the index of the expert with the smallest expected loss among those in A^t. Here, the expectations are taken over the algorithm's actions I_t (for a randomized algorithm), over the choice of $z_t \sim \mathcal{D}$, and over the generation of the awake sets A^t, when they are generated stochastically.

We restrict our study to the case where the awake sets are generated stochastically, possibly based on x_t, and focus on two separate scenarios, each requiring a different approach:

1. one where the awake sets are statistically independent of the losses, such that A^t is independent from z_t (Section 3);
2. one where the awake sets and the losses are statistically dependent (Section 4).

The notion of sleeping regret previously described was introduced by Kleinberg et al. (2008) for the first scenario, where the awake sets are independent of the losses. As we shall see, in the dependent case, this notion is no longer pertinent. Thus, we will generalize that expression and define a new notion of sleeping regret suitable for the dependent case.

3. Independent losses and awake sets

The independence between awake sets and losses is a crucial assumption in the study of online learning with sleeping experts by Kleinberg et al. (2008). Under this assumption, the authors presented an algorithm, called AUER, with tight theoretical guarantees. The algorithm is based on the classical Upper Confidence Bound (UCB) approach (Auer et al., 2002). It maintains a set of lower confidence bounds on the expected loss of each expert and, at each round, chooses the expert with the lowest confidence bound from the set of available experts. AUER is designed for the bandit setting, that is, when only the loss of the chosen expert is observed.

In this section, we present an extension of AUER to the feedback graph scenario, while assuming that the losses and awake sets are statistically independent. In this section, we also assume that the feedback graph G^t depends only on information up to time $t-1$; in particular, G^t does not depend on the losses $L(\xi_j, z_t)$ generated at time t. To be consistent with the notion of sleeping experts, we further assume that the graph G^t contains only vertices in A^t, although this does not affect the proofs. The pseudocode of our algorithm, AUER-N, which stands for AUER *with Neighbors*, is given in Algorithm 1.

ALGORITHM 1: AUER-N

Init: $Q_i(0) = 1$ for all $i \in [K]$.
for $t \geq 1$ **do**
  $\mathcal{S}_i(t-1) \leftarrow \sqrt{\frac{5\log(t)}{Q_i(t-1)}}, \forall i \in [K]$;
  Receive awake set $A^t \subseteq \mathcal{E}$;
  Receive graph G^t with out-neighbors N_i^t, $i \in [K]$;
  $I_t \leftarrow \operatorname{argmin}_{j \in A^t} \left\{ \widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \right\}$;
  for $j \in N_{I_t}^t$ **do**
    $Q_j(t) \leftarrow Q_j(t-1) + 1$;
    $\widehat{\mu}_{j,t} \leftarrow \frac{L(\xi_j, z_t)}{Q_j(t)} + \left(1 - \frac{1}{Q_j(t)}\right) \widehat{\mu}_j(t-1)$.
  end for
end for

The idea behind the design of AUER-N is to update the time-t estimate $\widehat{\mu}_j(t)$ of the expected loss of all experts with index j in the out-neighborhood $N_{I_t}^t$ of the chosen expert I_t at every round. These out-neighborhoods are determined by the time-t feedback graph G^t. As with AUER, the algorithm selects the awake expert with the smallest confidence bound.

We denote by $\mu_j = \mathbb{E}[L(\xi_j, z)]$ the expected loss of expert ξ_j, and assume an indexing consistent with the ranking of these losses: $\mu_1 < \mu_2 < \cdots < \mu_K$. For any $i < j$, we denote by $\Delta_{i,j} = \mu_j - \mu_i$ the decrease in expected loss from ξ_j to ξ_i and use the convention $\Delta_{j,j} = 0$ for any $j \in [K]$. We also denote by $T_j(t)$ the number of times expert ξ_j is selected by the algorithm up to time t, and by $Q_j(t)$ the number of times the loss of expert with index j is observed. The theorem below gives a bound on the sleeping regret of AUER-N in terms of $T_j(t)$ and $Q_j(t)$. These quantities are both algorithm-dependent, but, by definition, the ratio $\frac{T_j(t)}{Q_j(t)}$ is bounded by one and can be far smaller for dense graphs.

Theorem 1 *Assume that, for any $t \in [T]$, the feedback graph G^t depends only on information up to time $t-1$, and that the awake sets A^t are generated i.i.d., independently of the loss values $L(\xi_j, z_t)$, $j \in [K]$. Then, the sleeping regret of* AUER-N *after T rounds is upper bounded as follows:*

$$\mathrm{r}_T^{\text{SLEEP}}(\text{AUER-N}) \leq \sum_{j=2}^{K} \frac{40 \log T}{\Delta_{j-1,j}} \mathbb{E}\left[\max_{t \in [T]} \frac{T_j(t)}{Q_j(t)}\right] + 4 \sum_{j=2}^{K} \Delta_{1,j}.$$

The proof of this theorem and that of all other results are

[1] An alternative benchmark for the sleeping experts setting, not considered here, is the *specialist framework* of Freund et al. (1997).

given in the appendix. Since $\max_t \frac{T_j(t)}{Q_j(t)}$ is upper-bounded by one for all $j \in [K]$, the sleeping regret bound of AUER-N is always more favorable than that of AUER. In particular, if the number of times a learner choses an expert is equal to the number of times that expert was observed, that is, $T_j(t) = Q_j(t)$, for all j (as in the standard bandit setting), then we recover the sleeping regret bound of AUER (Kleinberg et al., 2008). On the other hand, in the full information setting, when $A_t = \mathcal{E}$ we have $Q_j(t) = t$ for all j and t, and $\sum_{j\in[K]} T_j(t) = t$. Thus, when the gap terms $\Delta_{j-1,j}$ are all comparable, the algorithm achieves an improvement in the regret bound by a factor of $\frac{1}{K}$, which, naturally, is a consequence of the K times more feedback received at each round, compared to the bandit setting.

We complement Theorem 1 by proving lower bounds showing that the regret of AUER-N is information-theoretically optimal, at least in the bandit scenario where the feedback graphs G^t only contain self-loops. In particular, we extend the lower bound of Kleinberg et al. (2008), which holds in the case of adversarially chosen awake sets, to the case where they are generated stochastically.

Theorem 2 *Under the assumptions of Theorem 1, assume that the graphs G^t only contain self-loops (bandit setting), and let ϕ be an online algorithm for the multi-armed bandit problem that never picks a suboptimal expert more than $o(T^\alpha)$ times over the course of T rounds for every $\alpha > 0$. Then, there exists a distribution according to which awake sets A^t are drawn i.i.d. and for which the sleeping regret of algorithm ϕ is at least $\Omega\left(\sum_{j=2}^K \frac{1}{\Delta_{j-1,j}} \log(T)\right)$ for large enough T depending on $(\mu_i)_{i\in[K]}$.*

4. Dependent losses and awake sets

In this section, we relax the assumption that the awake sets and losses are independent. We introduce a more general notion of sleeping regret, called *generalized sleeping regret* and argue that it is a more relevant notion of regret in the dependent setting than the standard sleeping regret defined in Kleinberg et al. (2008). We then present a sleeping experts algorithm based on the UCB algorithm of Auer et al. (2002), which we call UCB-SLG (UCB with SLeeping Graphs), that exploits feedback graphs and admits a favorable bound for the generalized sleeping regret. This will in turn pave the way for the application to the scenario of online abstention covered in Section 5, where the losses and awake sets are dependent and, where a natural notion of feedback graph over experts can be exploited.

4.1. Generalized sleeping regret

To see why the dependence between losses and awake sets invalidates the classical notion of sleeping regret, recall that the standard definition of sleeping regret (see Section 2) uses $\mathbb{E}[L(\xi, z)]$ to compare the expected loss of the chosen expert against the awake expert with the smallest expected loss. Then, consider the natural scenario where the input space $\mathcal{X}$ is the real line, and where an expert is awake only when $x > 0$, thereby making losses and awake sets both depend on x. Notice that the loss of this expert can only be (potentially) observed when $x > 0$. On the region of the space where $x < 0$, this loss might be arbitrarily large, but this has no effect on the loss incurred by any strategy that chooses this expert. Thus, the unconditional expectation, $\mathbb{E}[L(\xi, z)]$, used in Section 2 to define the notion of sleeping regret and adopted in Section 3 for the independent case, is no longer relevant. In the dependent setting, only the *conditional* expectation of the loss given that the expert is awake should be considered when defining regret.

Formally, let $\mathcal{A} = \{\mathtt{A}_1, \ldots, \mathtt{A}_p\}$ be the set of all possible awake sets, and let A be the random variable that generates the i.i.d. sequence of awake sets $\{A^t\}_{t=1}^T$ defined in Section 2. Then, the *generalized sleeping regret* $\mathrm{R}_T^{\text{SLEEP}}(\mathcal{B})$ of an algorithm $\mathcal{B}$ is defined as follows:

$$\mathrm{R}_T^{\text{SLEEP}}(\mathcal{B}) = \sum_{t=1}^T \sum_{k=1}^p p_k \, \mathbb{E}[L(\xi_{I_t}, z_t) - L(\xi_{i^*(k)}, z_t) | A^t = \mathtt{A}_k],$$

where $p_k = \mathbb{P}[A^t = \mathtt{A}_k]$ is the probability of the awake set $\mathtt{A}_k$, and $i^*(k) = \operatorname{argmin}_{i\in\mathtt{A}_k} \mathbb{E}[L(\xi_i, z) | A = \mathtt{A}_k]$.

When awake sets and losses are independent, the generalized sleeping regret coincides with the notion of sleeping regret of Section 2 since, for any t, the following holds:

$$\begin{aligned}
&\mathbb{E}[L(\xi_{\sigma(\mathtt{A}^t)}, z_t)] \\
&= \sum_{k=1}^p p_k \, \mathbb{E}[L(\xi_{\sigma(A^t)}, z_t) | A^t = \mathtt{A}_k] \\
&= \sum_{k=1}^p p_k \, \mathbb{E}[L(\xi_{\sigma(\mathtt{A}_k)}, z_t)] = \sum_{k=1}^p p_k \min_{i\in\mathtt{A}_k} \mathbb{E}[L(\xi_i, z_t)].
\end{aligned}$$

When all experts are awake at each round, the generalized sleeping regret matches the standard definition of regret.

4.2. The UCB-SLG algorithm

In view of the discussion above, the simple strategy adopted in AUER-N, that is, averaging over the time steps where an expert was awake and where its loss was observed, cannot work here, since this would lead to arbitrarily biased empirical estimates. This is typically the case where experts can be awake only for certain regions of the input space $\mathcal{X}$. Thus, our main idea for tackling the dependency between losses and awake sets is to use empirical estimates conditioned on the awake sets and decompose the problem into p subproblems, one per awake set $\mathtt{A}_k$, $k \in [p]$. Our UCB-SLG

ALGORITHM 2: The UCB-SLG algorithm.

Init: $O_{k,j}(0) = 1$ for all $j \in [K]$, and $k \in [p]$;
for $t \geq 1$ **do**
  $S_{k,j}(t-1) \leftarrow \sqrt{\frac{5\log(t)}{O_{k,j}(t-1)}}$ for all $k \in [p]$;
  Receive awake set $A^t \subseteq \mathcal{E}$;
  Let $k \in [p]$ be such that $A^t = \mathtt{A}_k$;
  Receive graph G_k with out-neighbors $N_{k,i}$, $i \in \mathtt{A}_k$;
  $I_t \leftarrow \operatorname{argmin}_{i \in \mathtt{A}_k} \widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1)$;
  for $j \in N_{k,I_t}$ **do**
    $O_{k,j}(t) \leftarrow O_{k,j}(t-1) + 1$;
    $\widehat{\nu}_{k,j}(t) \leftarrow \frac{L(\xi_j, z_t)}{O_{k,j}(t)} + \left(1 - \frac{1}{O_{k,j}(t)}\right) \widehat{\nu}_{k,j}(t-1)$.
  end for
end for

algorithm is based on this strategy. The algorithm estimates $\nu_{i,k} = \mathbb{E}[L(\xi_i, z)|A = \mathtt{A}_k]$ for each expert $\xi_i \in \mathcal{E}$ and each awake set $\mathtt{A}_k \in \mathcal{A}$. If the awake set at time t equals the awake set with index k, that is $A^t = \mathtt{A}_k$, then the algorithm chooses the expert with index $i \in \mathtt{A}_k$ with the lowest confidence bound of the empirical estimate of the conditional loss:

$$I_t \leftarrow \operatorname*{argmin}_{i \in \mathtt{A}_k} \widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1),$$

where $\widehat{\nu}_{k,i}(t-1)$ is the empirical estimate of $\nu_{k,i}$, and $S_{k,i}(t-1)$ is the corresponding slack (or confidence) term.

For simplicity, we present the algorithm and analysis under the assumption that each awake set admits a fixed feedback graph. Specifically, for each awake set $\mathtt{A}_k$, we define a fixed feedback graph $G_k = (\mathtt{A}_k, E_k)$, whose set of vertices is $\mathtt{A}_k$, the set of awake experts at that round, and whose set of edges E_k characterize the observability of the awake experts. Then, we assume that the graph at time t equals G_k, that is $G^t = G_k$, whenever $A^t = \mathtt{A}_k$. As a consequence, conditioned on $A^t = \mathtt{A}_k$, the graph G^t is independent of the losses at time t. We denote by $N_{k,i}$ the out-neighborhood of expert ξ_i on graph G_k. Given the above definitions, if at round t we have $A^t = \mathtt{A}_k$, then, UCB-SLG uses the losses of the experts in the out-neighborhood of N_{k,I_t} of graph G_k to update its empirical estimates. The pseudocode of UCB-SLG is given in Algorithm 2.

Our algorithm and analysis can be extended to the setting where the feedback graphs are not fixed per awake set and where instead we consider a sequence of time-varying graphs G_k^t per awake set $\mathtt{A}_k$ such that, conditioned on $A^t = \mathtt{A}_k$, the graphs G_k^t are independent of the losses at time t. We can extend the UCB-SLG algorithm in a natural way by updating all observed actions in the graph at each round. Moreover, the same analysis follows by considering the intersection over time of all the graphs in each region.

The regret guarantee of UCB-SLG depends on $\overline{\Delta}_{k,i} = \nu_{k,i} - \nu_{k,i^*(k)}$, the difference of the conditional losses for each awake set $k \in [p]$, as well as $\mathtt{B}_k = \{i \in \mathtt{A}_k \colon \nu_{k,i} = \nu_{k,i^*(k)}\}$, the set of arms whose conditional losses are equal to the optimal arm for the given awake set. Note that the learner incurs zero instantaneous regret when she chooses an expert with zero loss gap.

Our regret guarantee characterizes the benefit from the additional loss observations by partitioning the feedback graph G_k into cliques for each $k \in [p]$ and taking the minimum over all such possible clique coverings. Specifically, we define a *clique* of a directed graph G_k as a set of vertices $C \subset \mathtt{A}_k$ that are all neighbors with each other, that is such that, for all $i, j \in C$, we have $i \in N_{k,j}$ and $j \in N_{k,i}$. A clique covering $\mathcal{C}_k$ of graph G_k is defined as a set of cliques that satisfy $\cup_{C \in \mathcal{C}_k} = \mathtt{A}_k$. In the following theorem, the minimum is over all sets of all clique coverings $\mathcal{C}_k$ for each graph G_k with $k \in [p]$.

Theorem 3 *Assume that the sequence of awake sets A^t and loss values $L(\xi_j, z_t)$, $j \in [K]$, are generated jointly at time t, but i.i.d. over time. Assume further that each awake set $\mathtt{A}_k$ admits a fixed feedback graph G_k. Then, the generalized sleeping regret of the* UCB-SLG *algorithm after T rounds is bounded as follows:*

$$\mathrm{R}_T^{\text{SLEEP}}(\text{UCB-SLG}) = O\left(\sum_{k=1}^{p} p_k \min_{\mathcal{C}_k} \sum_{C \in \mathcal{C}_k} \frac{\max_{j \in C \setminus \mathtt{B}_k} \overline{\Delta}_{k,j}}{\min_{j \in C \setminus \mathtt{B}_k} (\overline{\Delta}_{k,j})^2} \log(T)\right).$$

The generalized sleeping regret is decomposed into a sum over the regret for each awake set $\mathtt{A}_k$ times the probability of that awake set. If the probability p_k of an awake set $\mathtt{A}_k$ is very small, then the bound on the regret for this awake set is given less weight. As one would expect for UCB-type algorithms, the regret is logarithmic in T for each awake set $\mathtt{A}_k$ but, unlike standard bounds, the loss gap is based on conditional expectations. This makes this bound not readily comparable to that of Theorem 1, where awake sets and losses are independent, even in the bandit setting.

Our regret bound shows the benefit of using feedback graphs. In particular, if the graphs are denser, then there are more ways to partition the graphs. Thus, the $\min_{\mathcal{C}_k}$ is over a larger set, thereby potentially decreasing the overall bound. In other words, if more losses are revealed at each round, the bound on the regret decreases accordingly. In the case where there is a single region, $p = 1$, with one fixed graph G^1, the bound reduces to the setting analyzed in Caron et al. (2012), and our regret bound for UCB-SLG matches that of the UCB-N algorithm therein.

5. Online learning with abstention

In this section, we apply the ideas introduced in Section 4 to the setting of online learning with abstention.

Online learning with abstention is a scenario recently introduced by Cortes et al. (2018), where a learner can elect to abstain from making a prediction at the price of a certain cost $c > 0$. When the learner abstains, she does not receive the label of the current input. The benefit of abstaining is that the loss incurred, c, is typically lower than the loss of incorrectly predicting a label.

While Cortes et al. (2018) cast the online abstention setting as an instance of online learning with feedback graphs, in this section, we cast the problem as an instance of online learning with both feedback graphs and sleeping experts. As we shall see, this choice provides us with a more meaningful and challenging benchmark for the learner, as well as an algorithm that achieves sublinear regret with respect to this benchmark. As a result, the algorithms we present achieve a more favorable empirical performance than those presented in Cortes et al. (2018), outperforming even an unrealistic full information algorithm designed only for the online abstention (but not feedback graph) setting.

We will adopt the notation of Cortes et al. (2018): let $r\colon \mathcal{X} \to \mathbb{R}$ denote an abstention function that dictates which examples to abstain on, and let $h\colon \mathcal{X} \mapsto \mathbb{R}$ denote a prediction function that determines the predicted labels of the examples. Let $\mathcal{E} = \{\xi_j = (h_j, r_j)\colon j \in [K]\} \subseteq \mathcal{H} \times \mathcal{R}$ denote a family of *experts* made up of pairs of a prediction function in $\mathcal{H}$ and an abstention function in $\mathcal{R}$.

The online abstention protocol is as follows. At each round $t \in [T]$, the learner receives an input point $x_t \in \mathcal{X}$ drawn i.i.d. according to the marginal distribution associated with $\mathcal{D}$, and chooses an index $I_t \in [K]$ corresponding to a pair $\xi_{I_t} = (h_{I_t}, r_{I_t})$. The learner determines whether to make a prediction based on the value of the abstention function, $r_{I_t}(x_t)$. If $r_{I_t}(x_t) \leq 0$, the learner abstains and incurs a fixed loss $c \in \mathbb{R}_+$. If $r_{I_t}(x_t) > 0$, the learner predicts, her prediction being $h_{I_t}(x_t)$. In this case (and only in this case), she receives a label $y_t \in \{\pm 1\}$, and incurs the prediction loss $\ell(y_t, h_{I_t}(x_t))$. One natural choice for an abstention function r associated with a prediction function h is a confidence-based function measuring the magnitude of h, that is $r(x) = |h(x)| - \gamma$ for some threshold, $\gamma \geq 0$ (e.g., (Bartlett & Wegkamp, 2008)). In the sequel, we focus on the binary classification problem where $\ell(y, h(x)) \in [0, 1]$ can be the 0/1 loss, $\mathbb{I}\{yh(x) \leq 0\}$,[2] or any of its bounded surrogates. The *abstention loss* of the pair $\xi = (h, r)$ on example $(x, y) \in \mathcal{X} \times \{\pm 1\}$ is defined as

$$L(\xi, z) = \ell(y, h(x))\mathbb{I}\{r(x) > 0\} + c\,\mathbb{I}\{r(x) \leq 0\}.$$

At first glance, the online abstention and sleeping expert settings appear to be different frameworks. Yet, we can cast the online abstention setting as a variant of the sleeping experts one by carefully defining an awake set that captures the loss incurred by any choice of arm. Naïvely, we can define an awake set to contain all the experts that are committed to a prediction. That is, if expert i is such that $r_i(x_t) > 0$, then we say that this expert is awake at time t, and if the expert is abstaining ($r_i(x_t) \leq 0$), then we say that this expert is asleep. However, we need to refine this definition of awake set, since we need to allow the algorithm to potentially abstain and incur a loss of c. Thus, we introduce an extra all-abstain expert, $\xi_0 = (h_0, r_0)$, whose abstention function r_0 is such that $r_0(x) \leq 0$ for all $x \in \mathcal{X}$, and define the awake set at time t to be made up of all the experts that are committed to a prediction, plus the all-abstain expert: $A^t = \{i \in [K]\colon r_i(x_t) > 0\} \cup \{0\}$.

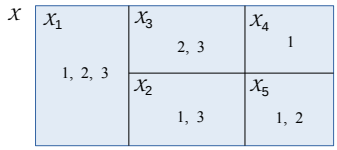

Figure 1: An online abstention scenario with $K = 3$ experts (excluding the all-abstain one). The input space $\mathcal{X}$ is partitioned into five regions $\mathcal{X}_1, \ldots, \mathcal{X}_5$, each one corresponding to a given combination of predicting experts on the inputs x belonging to that region. The set of predicting experts is indicated within each region. For instance, in region $\mathcal{X}_2$ we have, for all $x \in \mathcal{X}_2$, $r_1(x) > 0$, $r_2(x) \leq 0$, and $r_3(x) > 0$, while the prediction region of expert ξ_1, that is $\{x \in \mathcal{X}\colon r_1(x) > 0\}$, is $\mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_4 \cup \mathcal{X}_5$. The above gives rise to the $p = 5$ awake sets $\mathtt{A}_1 = \{0, 1, 2, 3\}$, $\mathtt{A}_2 = \{0, 1, 3\}$, $\mathtt{A}_3 = \{0, 2, 3\}$, $\mathtt{A}_4 = \{0, 1\}$, and $\mathtt{A}_5 = \{0, 1, 2\}$.

This definition in turn implies that $\mathtt{A}_k \in \mathcal{A}$ in the abstention setting is defined as the set of experts that are always awake together, plus the all-abstain expert. In other words, the input space $\mathcal{X}$ is partitioned into (disjoint) regions $\mathcal{X}_1, \ldots, \mathcal{X}_p$ each one corresponding to a given combination of predicting experts on that region. See Figure 1 for an illustration.

Without loss of generality, we assume that $|\mathtt{A}_k \setminus \{0\}| \geq 1$. If the labeled points $(x_1, y_1), \ldots, (x_T, y_T)$ are drawn i.i.d. according to some distribution $\mathcal{D}$ over $\mathcal{X} \times \{\pm 1\}$, then $p_k = \mathbb{P}(A^t = \mathtt{A}_k) = \mathbb{P}(x_t \in \mathcal{X}_k)$. Moreover, the sequence of awake sets $A^1, \ldots, A^T$ is an i.i.d. sequence and so is, for each $i \in [K]$, the sequence of losses $L(\xi_i, (x_1, y_1)), \ldots, L(\xi_i, (x_T, y_T))$. Yet, the random variables A^t and $L(\xi_i, (x_t, y_t))$ are not independent since they both depend on x_t.

Given the above connection between online abstention and sleeping experts, we can directly use the UCB-SLG algorithm presented in the previous section to obtain an algorithm for online learning with abstention. However, since we know that the loss of the all-abstain expert is always c, there is no need for a confidence interval for this expert. Thus, we present an algorithm called UCB-ABS, for UCB with ABStention, which is based on UCB-SLG, but also uses this

[2] Here, $\mathbb{I}\{\cdot\}$ denotes the indicator function.

ALGORITHM 3: UCB-ABS.

Init: $O_k(0) = 1$ for all $k \in [p]$;
for $t \geq 1$ **do**
  $S_k(t) \leftarrow \sqrt{\frac{5\log(t)}{O_k(t-1)}}$ for all $k \in [p]$;
  Receive $x_t \in \mathcal{X}$, and awake set $A^t \subseteq \mathcal{E}$;
  Let $k \in [p]$ be such that $A^t = \mathtt{A}_k$;
  Receive graph G_k with out-neighbors $N_{k,i}$, $i \in \mathtt{A}_k$;
  if $\min_{i \in \mathtt{A}_k \setminus \{0\}} \widehat{\nu}_{k,i}(t-1) - S_k(t-1) < c$ **then**
    $I_t \leftarrow \operatorname{argmin}_{i \in \mathtt{A}_k \setminus \{0\}} \widehat{\nu}_{k,i}(t-1)$;
    Reveal y_t;
    $O_k(t) \leftarrow O_k(t-1) + 1$;
    for $j \in \mathtt{A}_k \setminus \{0\}$ **do**
      $\widehat{\nu}_{k,j}(t) \leftarrow \frac{\ell(y_t, h_j(x_t))}{O_k(t)} + \left(1 - \frac{1}{O_k(t)}\right) \widehat{\nu}_{k,j}(t-1)$;
    end for
  else
    $I_t \leftarrow 0$;
  end if
end for

special property of the all-abstain expert.

Given the event $A^t = \mathtt{A}_k$, the UCB-ABS algorithm chooses the all-abstain expert if c is less than the smallest lowest confidence bound of $\nu_{k,i}$ over the experts $\xi_i \neq \xi_0$ in the awake set (i.e., all $i \in \mathtt{A}_k \setminus \{0\}$). If this is not the case, then it chooses the expert with index $i \in \mathtt{A}_k \setminus \{0\}$ having the smallest estimated conditional loss. In short, if $A^t = \mathtt{A}_k$, the UCB-ABS algorithm picks the expert with index:

$$I_t = \begin{cases} 0 & \text{if } c < \min_{i \in \mathtt{A}_k \setminus \{0\}} \widehat{\nu}_{k,i}(t-1) - S_k(t-1) \\ \operatorname{argmin}_{i \in \mathtt{A}_k \setminus \{0\}} \widehat{\nu}_{k,i}(t-1) & \text{otherwise.} \end{cases}$$

Algorithm 3 shows the pseudocode. Notice that, unlike the previous section, there is no need to maintain individual statistics for each expert here. In particular, the quantities $S_k(t)$ and $O_k(t)$ now refer to $\mathtt{A}_k$. This is due to the specific structure of the feedback graphs G_k, as explained below.

The losses revealed at each round depend on whether the chosen expert is the all-abstain expert. For each awake set $\mathtt{A}_k$, suppose that expert ξ_i is not the all-abstain expert. Then if this expert is chosen at time t, the true label y_t is revealed, so the loss of all experts is observed. On the other hand, if the all-abstain expert is chosen, then only the loss of the all-abstain expert is revealed. Thus, for each awake set $\mathtt{A}_k$, there is one fixed graph $G_k = (\mathtt{A}_k, E_k)$ whose out-neighborhoods are defined as follows: $N_{k,i} = \mathtt{A}_k$ if $i \neq 0$ and $N_{k,i} = \{0\}$ if $i = 0$. See Figure 2 for an illustration of the graph G_k. This feedback graph G_k is, in fact, the largest feedback graph per awake set that can be constructed in the abstention setting. Notice that after we condition over each set $A^t = \mathtt{A}_k$, this algorithm is almost running Follow-The-Leader in the full information setting on the experts of the awake set (excluding the all-abstain expert).

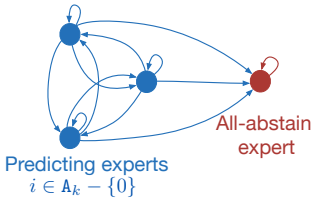

Figure 2: Illustration of graph G_k for the abstention setting. Here $\mathtt{A}_k$ contains four experts, including the all-abstain one.

The bound on the regret of UCB-ABS is again based on the conditional loss gaps $\overline{\Delta}_{k,i}$, clique coverings $\mathcal{C}_k$ and optimal arm sets $\mathtt{B}_k$ defined in the previous section.

Theorem 4 *The generalized sleeping standard regret of the* UCB-ABS *algorithm after T rounds is bounded as follows:*

$$\mathrm{R}_T^{\mathrm{SLEEP}}(\text{UCB-ABS}) = O\left(\sum_{k=1}^{p} p_k \min_{\mathcal{C}_k} \sum_{C \in \mathcal{C}_k} \frac{\max_{j \in C \setminus (\{0\} \cup \mathtt{B}_k)} \overline{\Delta}_{k,j}}{\min_{j \in C \setminus (\{0\} \cup \mathtt{B}_k)} (\overline{\Delta}_{k,j})^2} \log(T)\right).$$

Unlike the bound for UCB-SLG in Theorem 3, the maximum and minimum are not over all experts $j \in C \setminus \mathtt{B}_k$. Instead, they exclude the all-abstain expert. The proof of this theorem is similar to that of UCB-SLG except that it exploits the fact that no confidence interval is needed for the estimate of the all-abstain expert.

The regret bound above implies that the average cumulative loss of UCB-ABS will converge in $O((\log T)/T)$ to $\sum_{k=1}^{p} p_k \min_{i \in \mathtt{A}_k} \nu_{k,i}$. In contrast, previous algorithms for the abstention framework introduced in (Cortes et al., 2018) only prove bounds on the standard regret, which admit as benchmark $\min_{i \in [K]} \mu_i$. Observe that, by the super-additivity of the min operator and our treatment of the all-abstain expert, the following inequality holds:

$$\sum_{k=1}^{p} p_k \min_{i \in \mathtt{A}_k} \nu_{k,i} \leq \min_{i \in [K]} \sum_{k=1}^{p} p_k \nu_{k,i} = \min_{i \in [K]} \mu_i \,.$$

Thus, we expect UCB-ABS to outperform other abstention algorithms, which is corroborated in our experiments in Section 6.

The computational complexity of the algorithm depends on p since it keeps estimates of the conditional losses for each awake set in $\mathtt{A}_k$, $k \in [p]$. Since the awake sets are the intersections of the accepting regions, one can define abstention functions r_i such that the resulting number of awake sets p is not too large. For example, in the scenario where the hypothesis functions h_i perform well in complementary regions of the input space, we can define abstention functions whose non-abstention regions do not overlap in such a way that $p = K$. This is conceivable, for example, in a recommendation system setting, where regions correspond to

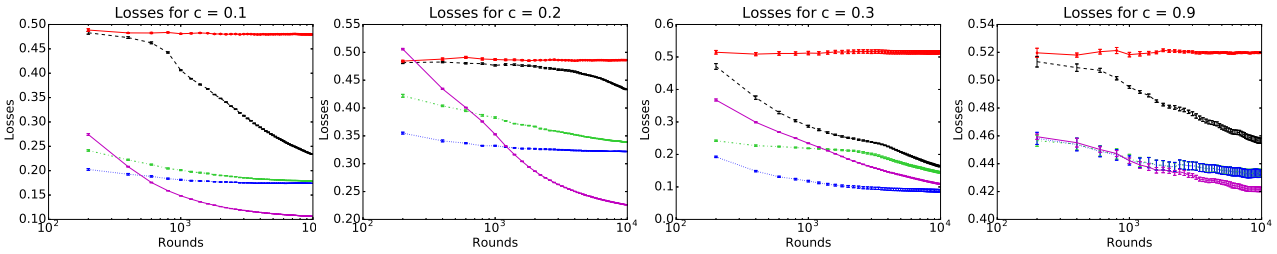

Figure 3: A graph of the averaged loss with standard deviations as a function of t (log scale). The algorithms we tested are UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. Starting from the left, the datasets are: `eye`, `HIGGS` , `skin`, and `covtype`.

general categories of an item and some hypothesis functions might be better at making recommendations within a certain category and hence should only be awake for that category.

6. Experiments

In this section, we present the results of several experiments for the online learning with abstention setting described in Section 5. These experiments demonstrate that UCB-ABS admits a strong empirical performance.

We compare UCB-ABS to several baselines, including the algorithms UCB-GT, UCB-NT, and FS of Cortes et al. (2018), as well as standard UCB (Auer et al., 2002). FS is the ideal comparator that picks the expert with the smallest empirical mean, but has the advantage that the losses of all the experts are revealed at each round. Thus FS and UCB lie at the two ends of a spectrum. On one end, the FS algorithm has access to the full loss information at each round, at the other end, the UCB algorithm only sees the loss of the expert chosen. In between these two extremes, the losses revealed to UCB-GT, UCB-NT, and UCB-ABS depend on whether the chosen expert for each algorithm abstains or predicts. That is, if the chosen expert predicts at time t, the true label y_t is revealed and hence the losses of all the experts are observed, while if it abstains at time t, only the loss of the abstaining experts, which is simply equal to the abstention cost c, is revealed. It is important to note that UCB-ABS uses all the revealed expert losses to update its empirical estimates at each round, while UCB-GT and UCB-NT only use a subset of the revealed losses at each round, since these latter algorithms can only make updates based on information up to time $t-1$. On the other hand, FS relies on full information no matter what, and is therefore relying on informational assumptions which are clearly outside the abstention setting.

For ease of comparison, in our experiments, we adopted the same setup and used the same datasets as Cortes et al. (2018). That is, the predictions functions h are random hyperplanes centered at the origin with normal vectors drawn randomly from the Gaussian distribution $\mathcal{N}(0,1)^d$ where d is the feature dimension, and the abstention functions r are concentric annuli centered at the origin with radii in $(0, \frac{\sqrt{d}}{20}, \frac{2\sqrt{d}}{20}, \ldots, \sqrt{d})$. We tested abstention costs c in $\{0.05, 0.1, 0.15, \ldots, 0.9\}$. We used the `CIFAR` dataset from Krizhevsky et al. (2009), where we extracted the first twenty-five principal components, and used eight UCI datasets: `HIGGS`, `phishing`, `ijcnn`, `covtype`, `eye`, `skin`, `cod-rna`, and `guide`. The loss of each algorithm was calculated as follows: First we fixed the set of experts and averaged the results over five random draws of the data, and then let the experts vary and averaged the results over five random draws of the experts.

Figure 3 shows the averaged abstention loss $L(\cdot)/t$ with standard deviations for the different abstention costs. In Appendix E, we show the plots of all the datasets we tested, where the same patterns recur. These experiments show that UCB-ABS outperforms UCB-NT and UCB on all datasets and it attains a better averaged loss than that of UCB-GT on most datasets. Remarkably, on some datasets UCB-ABS even outperforms FS, which is an unrealistic baseline that clearly violates the rules of the abstention setup in that this algorithm receives all loss information at each round. This algorithm was used for its ideal performance in Cortes et al. (2018). Thus, thanks to a generalized notion of sleeping regret and casting the abstention problem as an instance of the sleeping experts framework, we obtain both theoretical and empirical improvements. In addition to the experiments above, we tested the effects of increasing the number of abstention and prediction functions. We also present plots for the fraction of points each algorithm abstains on – see Appendix E.

7. Conclusion

We presented a comprehensive analysis of online learning with sleeping experts and feedback graphs, combining two lines of existing work that are closely related but have so far not been considered together. We presented both algorithmic solutions and theoretical analysis, and we also adapted our ideas to the online abstention problem, with extensive experiments showing that our adaptation outperforms existing solutions. While our experiments focused on binary classification, they can be directly extended to multiclass classification and regression problems.

References

Alon, Noga, Cesa-Bianchi, Nicolò, Gentile, Claudio, and Mansour, Yishay. From bandits to experts: A tale of domination and independence. In *NIPS*, 2013.

Alon, Noga, Cesa-Bianchi, Nicolò, Dekel, Ofer, and Koren, Tomer. Online learning with feedback graphs: Beyond bandits. In *JMLR*, pp. 23–35, 2015.

Auer, Peter, Cesa-Bianchi, Nicolò, and Fischer, Paul. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, 2002.

Bartlett, Peter and Wegkamp, Marten. Classification with a reject option using a hinge loss. *JMLR*, pp. 291–307, 2008.

Buccapatnam, Swapna, Eryilmaz, Atilla, and Shroff, Ness B. Stochastic bandits with side observations on networks. In *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '14, pp. 289–300. ACM, 2014.

Caron, Stephane, Kveton, Branislav, Lelarge, Marc, and Bhagat, Smriti. Leveraging side observations in stochastic bandits. In *UAI*, 2012.

Cesa-Bianchi, Nicolò and Lugosi, Gábor. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA, 2006.

Cohen, Alon, Hazan, Tamir, and Koren, Tomer. Online learning with feedback graphs without the graphs. *arXiv preprint arXiv:1605.07018*, 2016.

Cortes, Corinna, DeSalvo, Giulia, Gentile, Claudio, Mohri, Mehryar, and Yang, Scott. Online learning with abstention. In *ICML*, 2018.

Freund, Yoav, Schapire, Robert E, Singer, Yoram, and Warmuth, Manfred K. Using and combining predictors that specialize. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pp. 334–343. ACM, 1997.

Kanade, Varun and Steinke, Thomas. Learning hurdles for sleeping experts. *ACM Transactions on Computation Theory (TOCT)*, 6(3):11, 2014.

Kanade, Varun, McMahan, H Brendan, and Bryan, Brent. Sleeping experts and bandits with stochastic action availability and adversarial rewards. In *Artificial Intelligence and Statistics*, pp. 272–279, 2009.

Kleinberg, Robert, Niculescu-Mizil, Alexandru, and Sharma, Yogeshwer. Regret bounds for sleeping experts and bandits. *Machine Learning*, 2008.

Kocák, Tomáš, Neu, Gergely, Valko, Michal, and Munos, Rémi. Efficient learning by implicit exploration in bandit problems with side observations. In *NIPS*, pp. 613–621, 2014.

Kocák, Tomáš, Neu, Gergely, and Valko, Michal. Online learning with erdős-rényi side-observation graphs. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 339–346. AUAI Press, 2016.

Krizhevsky, Alex, Nair, Vinod, and Hinton, Geoffrey. Cifar-10 (canadian institute for advanced research), 2009. URL http://www.cs.toronto.edu/~kriz/cifar.html.

Littlestone, Nick and Warmuth, Manfred K. The weighted majority algorithm. *Information and computation*, 108 (2):212–261, 1994.

Liu, Fang, Buccapatnam, Swapna, and Shroff, Ness. Information directed sampling for stochastic bandits with graph feedback. In *32nd AAAI Conference on Artificial Intelligence*, 2018.

Mannor, Shie and Shamir, Ohad. From bandits to experts: On the value of side-observations. In *NIPS*, pp. 291–307, 2011.

Tossou, Aristide, Dimitrakakis, Christos, and Dubhashi, Devdatt. Thompson sampling for stochastic bandits with graph feedback. In *31st AAAI Conference on Artificial Intelligence*, 2017.

Wu, Yifan, György, András, and Szepesvari, Csaba. Online learning with gaussian payoffs and side observations. In *Advances in Neural Information Processing Systems 28*, pp. 1360–1368. Curran Associates, Inc., 2015.

Yun, Donggyu, Proutiere, Alexandre, Ahn, Sumyeong, Shin, Jinwoo, and Yi, Yung. Multi-armed bandit with additional observations. *Proc. ACM Meas. Anal. Comput. Syst.*, 2 (1):13:1–13:22, 2018.

A. AUER-N

In this section, we present the proof of the regret guarantee for AUER-N. For the sake of this analysis, we are in fact assuming that the awake sets A^t are generated *arbitrarily* before learning starts. This implies the claimed result when the A^t's are generated i.i.d. according to an arbitrary distribution over $\mathcal{A}$, independent of the losses. We use $\mathbb{I}\{\cdot\}$ to denote the indicator function.

We start off with the following technical lemma.

Lemma 1 *Assume the following ordering for μ_j, $j \in [K]$: $\mu_1 < \mu_2 < \cdots < \mu_K$ and, for $i < j$, let $\Delta_{i,j} = \mu_j - \mu_i$. Then, for any $F_j > 0$, $j \in [K]$, the following inequality holds:*

$$\sum_{1 \le i < j \le K} F_j \frac{\Delta_{i,i+1}}{\Delta_{i,j}^2} \le 2 \sum_{j=2}^{K} \frac{F_j}{\Delta_{j-1,j}} \,.$$

Proof. The result and the proof are extensions of Lemma 3 of Kleinberg et al. (2008) to inequalities augmented with factors F_j. We will use the following equality which, by definition of the Lebesgue integral, holds for any non-negative function f: $\mathbb{E}[f(X)] = \int_0^{+\infty} \mathbb{P}[f(x) \ge t]\, dt$. Thus, considering, in particular, the uniform probability over $\{2, \ldots, K\}$, we can write:

$$\begin{aligned}
\frac{1}{K-1} \sum_{j=2}^{K} \frac{\mathbb{I}\{j > i\} F_j}{\Delta_{i,j}^2} &= \int_0^{+\infty} \mathbb{P}\left[\frac{\mathbb{I}\{j > i\}\, F_j}{\Delta_{i,j}^2} \ge t\right] dt = \int_0^{+\infty} \mathbb{P}\left[\frac{\mathbb{I}\{j > i\}\, F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge \sqrt{t}\right] dt \\
&= 2 \int_0^{+\infty} \mathbb{P}\left[\frac{\mathbb{I}\{j > i\}\, F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge u\right] u\, du \qquad (u = \sqrt{t}) \\
&= \frac{2}{K-1} \int_0^{+\infty} \sum_{j=2}^{K} \mathbb{I}\{\frac{\mathbb{I}\{j > i\}\, F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge t\}\, t\, dt.
\end{aligned}$$

In view of the above, we have

$$\begin{aligned}
\sum_{1 \le i < j \le K} F_j \frac{\Delta_{i,i+1}}{\Delta_{i,j}^2} &= \sum_{i=1}^{K-1} \Delta_{i,i+1} \sum_{j:\, j>i} \frac{F_j}{\Delta_{i,j}^2} \\
&= 2 \sum_{i=1}^{K-1} \Delta_{i,i+1} \int_0^{+\infty} \sum_{j=2}^{K} \mathbb{I}\{\frac{\mathbb{I}\{j > i\}\, F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge t\}\, t\, dt \\
&= 2 \int_0^{+\infty} \sum_{i=1}^{K-1} \Delta_{i,i+1} \sum_{j=2}^{K} \mathbb{I}\{\frac{\mathbb{I}\{j > i\}\, F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge t\}\, t\, dt \\
&= 2 \int_0^{+\infty} \sum_{1 \le i < j \le K} \Delta_{i,i+1}\, \mathbb{I}\{\frac{F_j^{\frac{1}{2}}}{\Delta_{i,j}} \ge t\} t\, dt.
\end{aligned}$$

Now, for any $j \in \{2, \ldots, K\}$ and $t > 0$, define $i_t(j)$ by

$$i_t(j) = \operatorname{argmin} \left\{ i \in [K] \colon i \le j, \Delta_{i,j} \le \frac{F_j^{\frac{1}{2}}}{t} \right\}.$$

The index $i_t(j)$ is well defined since for $i = j$, $\Delta_{j,j} = 0$ is upper bounded by $\frac{F_j^{\frac{1}{2}}}{t}$. By definition of $i_t(j)$, we can write

$$\begin{aligned}
\sum_{1\le i<j\le K} F_j \frac{\Delta_{i,i+1}}{\Delta_{i,j}^2} = 2\int_0^{+\infty} \sum_{j=2}^{K} \sum_{i=i_t(j)}^{j-1} \Delta_{i,i+1}\, t\, dt &= 2\int_0^{+\infty} \sum_{j=2}^{K} \Delta_{i_t(j),j}\, t\, dt \\
&= 2\sum_{j=2}^{K} \int_0^{+\infty} \Delta_{i_t(j),j}\, t\, dt \\
&= 2\sum_{j=2}^{K} \int_0^{\frac{F_j^{\frac{1}{2}}}{\Delta_{j-1,j}}} \Delta_{i_t(j),j}\, t\, dt \qquad \left(\text{for } t \ge \frac{F_j^{\frac{1}{2}}}{\Delta_{j-1,j}}, i_t(j) = j\right) \\
&\le 2\sum_{j=2}^{K} \int_0^{\frac{F_j^{\frac{1}{2}}}{\Delta_{j-1,j}}} F_j^{\frac{1}{2}}\, dt \\
&= 2\sum_{j=2}^{K} \frac{F_j}{\Delta_{j-1,j}} \qquad \left(\text{by def. } \Delta_{i_t(j),j} \le \frac{F_j^{\frac{1}{2}}}{t}\right),
\end{aligned}$$

which completes the proof. □

With the above lemma handy, we are ready to prove Theorem 1.

Proof. [Theorem 1] For any $i, j \in [K]$, $i < j$, let $M_{i,j}$ denote the number of times expert ξ_j is selected by the algorithm, while some expert ξ_k with $k \in [i]$ could have been selected (because it was awake), where $[i] = \{1, \dots, i\}$. By definition, $(M_{i,j} - M_{i-1,j})$ is then the number of times expert ξ_j,is selected by the algorithm, while expert ξ_i, $i < j$, could have been selected. Then, using the convention $\Delta_{j,j} = 0$ and $M_{0,j} = 0$ for any $j \in [K]$, the sleeping regret of the algorithm can be expressed as follows:

$$\mathrm{r}_T^{\text{SLEEP}}(\text{AUER-N}) = \mathbb{E}\left[\sum_{1\le i<j\le K} (M_{i,j} - M_{i-1,j})\Delta_{i,j}\right] = \sum_{j=2}^{K}\sum_{i=1}^{j-1} \mathbb{E}[M_{i,j}][\Delta_{i,j} - \Delta_{i+1,j}]\,. \tag{1}$$

Thus, to bound the sleeping regret, it suffices to bound $\mathbb{E}[M_{i,j}]$ for $1 \le i < j \le K$. This expectation can be rewritten as follows

$$\mathbb{E}[M_{i,j}] = \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t = j\}\, \mathbb{I}\{A^t \cap [i] \neq \emptyset\}\right], \tag{2}$$

where A^t denotes set of experts awake at time t. For any random variable $\sigma_{i,j} \in [T]$, the above expression can be split into two sums:

$$(2) = \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t = j\}\, \mathbb{I}\{A^t \cap [i] \neq \emptyset\}\, \mathbb{I}\{T_j(t-1) < \sigma_{i,j}\}\right] + \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t = j\}\, \mathbb{I}\{A^t \cap [i] \neq \emptyset\}\, \mathbb{I}\{T_j(t-1) \ge \sigma_{i,j}\}\right].$$

Now, define $T^* = \max\left\{t \in [T] \colon \mathbb{I}\{T_j(t-1) < \sigma_{i,j}\} \neq 0\right\}$. Then, by definition, we have

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{I}\{I_t = j\}\mathbb{I}\{A^t \cap [i] \neq \emptyset\}\mathbb{I}\{T_j(t-1) < \sigma_{i,j}\} &= \sum_{t=1}^{T^*} \mathbb{I}\{I_t = j\}\mathbb{I}\{A^t \cap [i] \neq \emptyset\}\mathbb{I}\{T_j(t-1) < \sigma_{i,j}\} \\
&\le \sum_{t=1}^{T^*} \mathbb{I}\{I_t = j\} \\
&= T_j(T^*) \le T_j(T^* - 1) + 1 \le \sigma_{i,j}\,.
\end{aligned}$$

This shows that, for any $\sigma_{i,j} \in [T]$,

$$(2) \le \mathbb{E}\left[\sigma_{i,j} + \sum_{t=\sigma_{i,j}+1}^{T} \mathbb{I}\{I_t = j\}\mathbb{I}\{A^t \cap [i] \neq \emptyset\}\mathbb{I}\{T_j(t-1) \ge \sigma_{i,j}\}\right].$$

If expert j is selected at time t, that is $I_t = j$, then it must have the lowest confidence bound: $\widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \leq \widehat{\mu}_k(t-1) - \mathcal{S}_k(t-1)$ for all $k \in A^t$. Let $k^* = \operatorname{argmin}_{k \in A^t \cap [i]} \widehat{\mu}_k(t-1) - \mathcal{S}_k(t-1)$, then

$$\begin{aligned}\mathbb{E}[\mathbb{I}\{I_t = j\}\mathbb{I}\{A^t \cap [i] \neq \emptyset\}\mathbb{I}\{T_j(t-1) \geq \sigma_{i,j}\}] \\ \leq \mathbb{P}\left(\widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \leq \widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1), A^t \cap [i] \neq \emptyset, T_j(t-1) \geq \sigma_{i,j}\right) ,\end{aligned} \tag{3}$$

Next, the first event in the probability can be expressed as follows:

$$\begin{aligned}&\widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1) - \widehat{\mu}_j(t-1) + \mathcal{S}_j(t-1) \geq 0 \\ \Leftrightarrow &\widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1) - \widehat{\mu}_j(t-1) + \mathcal{S}_j(t-1) - \mathcal{S}_j(t-1) - \mu_{k^*} + \mu_{k^*} - \mu_j + \mu_j - \mu_i + \mu_i + \mathcal{S}_j(t-1) \geq 0 \\ \Leftrightarrow &\Big[-\mu_{k^*} + \widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1)\Big] + \Big[\mu_j - \widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1)\Big] + \Big[(\mu_{k^*} - \mu_i) - (\mu_j - \mu_i) + 2\mathcal{S}_j(t-1)\Big] \geq 0 .\end{aligned}$$

Thus, for that event to hold, at least one of these three terms must be non-negative. Moreover, if one is non-positive, at least one of the other two is non-negative.

Choose random variable $\sigma_{i,j}$ as follows: $\sigma_{i,j} = \frac{20 \log T}{(\mu_j - \mu_i)^2} \max_{t \in [T]} \frac{T_j(t-1)}{Q_j(t-1)}$. Then, the second event in the probability, $T_j(t-1) \geq \sigma_{i,j}$, implies

$$T_j(t-1) \geq \frac{20 \log T}{(\mu_j - \mu_i)^2} \frac{T_j(t-1)}{Q_j(t-1)} \Rightarrow (\mu_j - \mu_i)^2 \geq \frac{20 \log t}{Q_j(t-1)} \Leftrightarrow \mu_j - \mu_i \geq \sqrt{\frac{20 \log t}{Q_j(t-1)}} .$$

In view of that, when the second event $T_j(t-1) \geq \sigma_{i,j}$ holds, we have

$$\begin{aligned}(\mu_{k^*} - \mu_i) - (\mu_j - \mu_i) + 2\mathcal{S}_j(t-1) &\leq -(\mu_j - \mu_i) + 2\mathcal{S}_{j,t-1} && \text{(def. of } \mu_{k^*}) \\ &= -(\mu_j - \mu_i) + \sqrt{\frac{20 \log t}{Q_j(t-1)}} \leq 0 .\end{aligned}$$

This shows that the third term above is then non-positive and that at least one of the first two is non-negative. Thus, under the above choice of $\sigma_{i,j}$, the following inequality holds:

$$\begin{aligned}&\mathbb{P}\left(\widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \leq \widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1), A^t \cap [i] \neq \emptyset, T_j(t-1) \geq \sigma_{i,j}\right) \\ &\qquad \leq \mathbb{P}\left(-\mu_{k^*} + \widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1) \geq 0\right) + \mathbb{P}\left(\mu_j - \widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \geq 0\right).\end{aligned} \tag{4}$$

Now, since both the feedback graph G^t and the algorithm's action I_t only depend on information up to time $(t-1)$, it is straightforward to see that, for any $j \in [K]$, the sequence of random variables $L(\xi_j, z_{s_1}), L(\xi_j, z_{s_2}), \ldots$, are i.i.d., and distributed as $L(\xi_j, z_1)$, where s_k is the stopping time $s_k = \min\{t \colon Q_j(t) = k\}$. Using a standard Hoeffding bound, this allows us to bound the second probability in (4) as follows:

$$\begin{aligned}&\mathbb{P}\left(\mu_j - \widehat{\mu}_j(t-1) - \mathcal{S}_j(t-1) \geq 0\right) \\ &\quad = \mathbb{P}\left(\mu_j - \frac{1}{Q(t)} \sum_{s=1}^{t} L(\xi_j, z_s)\mathbb{I}\{j \in N_{I_s}^s\} - \sqrt{\frac{5 \log t}{Q_j(t)}} \geq 0\right) \\ &\quad \leq \sum_{n=1}^{t} \mathbb{P}\left(\mu_j - \frac{1}{n} \sum_{s=1}^{t} L(\xi_j, z_s)\mathbb{I}\{j \in N_{I_s}^t\} - \sqrt{\frac{5 \log t}{n}} \geq 0 \wedge Q_j(t) = n\right) \\ &\quad = \sum_{n=1}^{t} \mathbb{P}\left(\mu_j - \frac{1}{n} \sum_{i=1}^{n} L(\xi_j, z_s) - \sqrt{\frac{5 \log t}{n}} \geq 0\right) \\ &\quad \leq \sum_{n=1}^{t} \frac{1}{t^5} = \frac{1}{t^4} .\end{aligned}$$

The other probability in (4), i.e., $\mathbb{P}(-\mu_{k^*} + \widehat{\mu}_{k^*}(t-1) - \mathcal{S}_{k^*}(t-1) \geq 0)$, can be bounded in a similar way, thereby resulting in the following upper bound:

$$\mathbb{E}[M_{i,j}] \leq \frac{20 \log T}{(\mu_j - \mu_i)^2} \mathbb{E}\left[\max_{t \in [T]} \frac{T_j(t-1)}{Q_j(t-1)}\right] + 2 \sum_{t=1}^{T} \frac{1}{t^4} \leq \frac{20 \log T}{(\mu_j - \mu_i)^2} \mathbb{E}\left[\max_{t \in [T]} \frac{T_j(t)}{Q_j(t)}\right] + 4 \,.$$

Plugging in the right-hand side of this inequality in (1) to upper-bound $\mathbb{E}[M_{i,j}]$, and using Lemma 1 with $F_j = \mathbb{E}\left[\max_{t \in [T]} \frac{T_j(t)}{Q_j(t)}\right]$ completes the proof. □

B. Lower bound on sleeping regret

This section provides a proof of the lower bound in Theorem 2. The proof of this result follows from extending the arguments in Kleinberg et al. (2008).

Proof. [Theorem 2] We first restate Lemma 11 in Kleinberg et al. (2008).

Lemma 2 (Lemma 11, Kleinberg et al. (2008)) *Suppose we are given two numbers $\mu_1 > \mu_2$, both lying in an interval $[a, b]$ such that $0 < a < b < 1$, and suppose we are given any online algorithm ϕ for the multi-armed bandit problem with two experts which never picks the worse expert more than $o(T^\alpha)$ times for every $\alpha > 0$. Then there is an input instance in the stochastic rewards model, with two experts L and R whose payoff distributions are Bernoulli random variables with means μ_1 and μ_2 or vice-versa, such that for large enough T depending on a, b, μ_1, and μ_2, the regret of algorithm ϕ is $\Omega\left(\frac{\log(T)(\mu_1-\mu_2)}{KL(\mu_2||\mu_1)}\right)$, where the constant inside the $\Omega(\cdot)$ is at least $\frac{1}{2}$.*

Assume that the losses are Bernoulli random variables and that the means $\{\mu_j\}_{j=1}^K$ are bounded away from 0 and 1. Let $A^t \subset [K]$ be the awake set at time t, and suppose that $A^t \sim U(\{2j-1, 2j\}_{j=1}^{K/2})$ independent of the distribution of the losses. For each awake set A and $t \in [T]$, let $s(A, t) \in [T]$ be the time step in which the awake set A occurred for the t-th time. Then we can write for the expected sleeping regret of any algorithm:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t} - \mu_{\sigma(A^t)}\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{K/2} \mathbb{I}\{A^t = \{2j-1, 2j\}\}(\mu_{I_t} - \mu_{\sigma(A^t)})\right] \\
&= \sum_{j=1}^{K/2} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}\{A^t = \{2j-1, 2j\}\}(\mu_{I_t} - \mu_{\sigma(A^t)})\right] \\
&= \sum_{j=1}^{K/2} \mathbb{E}\left[\sum_{t=1}^{\mathbb{E}[\sum_{t=1}^{T} \mathbb{I}\{A^t=\{2j-1,2j\}\}]} (\mu_{I_{s(\{2j-1,2j\},t)}} - \mu_{\sigma(A^t)})\right] \\
&= \sum_{j=1}^{K/2} \mathbb{E}\left[\sum_{t=1}^{\frac{2T}{K}} (\mu_{I_{s(\{2j-1,2j\},t)}} - \mu_{\sigma(A^t)})\right] \\
&\geq \sum_{j=1}^{K/2} \Omega\left(\frac{\log(2T/K)(\mu_{2j-1} - \mu_{2j})}{\mathrm{KL}(\mu_{2j-1}||\mu_{2j})}\right),
\end{aligned}$$

where the second to last equality follows from Wald's equation, and the inequality follows from applying Lemma 2 to each awake set which is effectively a separate two-armed bandit problem.

Now, since we assume that the means are bounded between a and b, we can upper bound the KL divergence terms as follows:

$$\mathrm{KL}(\mu_{2j-1}||\mu_{2j}) \leq \frac{(\mu_{2j-1} - \mu_{2j})^2}{\mu_{2j}(1-\mu_{2j})} \leq \frac{(\mu_{2j-1} - \mu_{2j})^2}{\min_{i=1}^{K} \mu_i(1-\mu_i)}.$$

Thus, we can write

$$\sum_{j=1}^{K/2} \Omega\left(\frac{\log(2T/K)(\mu_{2j-1} - \mu_{2j})}{\mathrm{KL}(\mu_{2j-1}||\mu_{2j})}\right) \geq \sum_{j=1}^{K/2} \Omega\left(\frac{\log(2T/K)}{\mu_{2j-1} - \mu_{2j}}\right).$$

Similarly, if we consider $A^t \sim U(\{2j, 2j+1\}_{j=1}^{K/2-1})$, then the expected sleeping regret of any algorithm is lower bounded by:

$$\mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t} - \mu_{\sigma(A^t)}\right] \geq \sum_{j=1}^{K/2-1} \Omega\left(\frac{\log(2T/K)}{\mu_{2j} - \mu_{2j+1}}\right).$$

Thus, if consider an awake set distribution $A^t \sim U(\{2j-1, 2j\}_{j=1}^{K/2})$ and $A^t \sim U(\{2j, 2j+1\}_{j=1}^{K/2-1})$ each with probability 1/2, then the expected sleeping regret of any algorithm is lower bounded by:

$$\mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t} - \mu_{\sigma(A^t)}\right] \geq \sum_{j=2}^{K} \Omega\left(\frac{\log(2T/K)}{\mu_{j-1} - \mu_j}\right).$$

□

C. UCB-SLG

In this section, we prove Theorem 3.

Proof. [Theorem 3] To simplify the notation, throughout this proof, we replace $L(\cdot, z_t)$ by $L_t(\cdot)$, $\mathbb{E}[\cdot|A_t = \mathtt{A}_k]$ by $\mathbb{E}[\cdot|\mathtt{A}_k]$ and $\nu_{k,i^*(k)}$ by $\nu_{i^*(k)}$. We first decompose the regret in terms of the awake sets $\mathtt{A}_1, \dots, \mathtt{A}_p$:

$$\mathrm{R}_T^{\text{SLEEP}}(\text{UCB-SLG}) = \sum_{t=1}^{T}\sum_{k=1}^{p} p_k \,\mathbb{E}\left[(L_t(\xi_{I_t}) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k\right] = \sum_{k=1}^{p} p_k \mathrm{R}_{T,k} \;,$$

where $\mathrm{R}_{T,k} = \sum_{t=1}^{T} \mathbb{E}[(L_t(\xi_{I_t}) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k]$ can be interpreted as the regret for region k at time T. Thus, we can focus on bounding $\mathrm{R}_{T,k}$ for each $k \in [p]$.

Fix $k \in [p]$. Observe that we can disregard any term in $\mathrm{R}_{T,k}$ where the conditional expectation of the chosen expert is less than that of the best expert, $\nu_{k,I_t} \le \nu_{i^*(k)}$, and bound that by zero:

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{\nu_{k,I_t} \le \nu_{i^*(k)}\}(L_t(\xi_{I_t}) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k\right] &\le \sum_{t=1}^{T}\sum_{i=1}^{K} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} \le \nu_{i^*(k)}\}(L_t(\xi_i) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k\right] \\ &\le \sum_{t=1}^{T}\sum_{i=1}^{K} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} \le \nu_{i^*(k)}\}|\mathtt{A}_k\right](\nu_{k,i} - \nu_{i^*(k)}) \le 0 \;,\end{aligned}$$

where in the second to last inequality, we used the fact that $(L_t(\xi_i) - L_t(\xi_{i^*(k)}))$ and $\mathbb{I}\{I_t = i\}$ are conditionally independent given $\mathtt{A}_k$. Thus, $\mathrm{R}_{T,k}$ can be upper bounded by terms where the conditional expectation of the chosen expert is greater than that of the best expert, $\nu_{k,I_t} > \nu_{i^*(k)}$:

$$\begin{aligned}\mathrm{R}_{T,k} &\le \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{\nu_{k,I_t} > \nu_{i^*(k)}\}(L_t(\xi_{I_t}) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k\right] \\ &= \sum_{t=1}^{T}\sum_{i\in\mathtt{A}_k} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}(L_t(\xi_i) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k\right] && (I_t \text{ must be in } \mathtt{A}_k) \\ &= \sum_{t=1}^{T}\sum_{i\in\mathtt{A}_k} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\right](\nu_{k,i} - \nu_{i^*(k)}), && (\text{cond. indep.}) \\ &= \sum_{t=1}^{T}\sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\right]\overline{\Delta}_{k,i} && (\text{def. of } \mathtt{B}_k \text{ and } \overline{\Delta}_{k,i}) \\ &= \sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}(r^1_{T,k,i} + r^2_{T,k,i}) \;,\end{aligned}$$

with $r^1_{T,k,i}$ and $r^2_{T,k,i}$ defined by

$$\begin{aligned}r^1_{T,k,i} &= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}\mathbb{I}\{O_{k,i}(t-1) < s_i\}|\mathtt{A}_k\right] \\ r^2_{T,k,i} &= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}\mathbb{I}\{O_{k,i}(t-1) \ge s_i\}|\mathtt{A}_k\right] \;,\end{aligned}$$

where s_i is a parameter whose value will be selected later. Since the event $I_t = i$ implies in particular the inequality $\widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1) \le \widehat{\nu}_{k,i^*(k)}(t-1) - S_{k,i^*(k)}(t-1)$, we have

$$r^2_{T,k,i} \le \sum_{t=1}^{T} \mathbb{P}\left[\widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1) \le \widehat{\nu}_{k,i^*(k)}(t-1) - S_{k,i^*(k)}(t-1) \wedge \nu_{k,i} > \nu_{i^*(k)} \wedge O_{k,i}(t-1) \ge s_i \middle| \mathtt{A}_k\right] .$$

The inequality defining the first event in this probability can be decomposed as follows:

$$\begin{aligned}
&\widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1) \le \widehat{\nu}_{k,i^*(k)}(t-1) - S_{k,i^*(k)}(t-1) \\
\Leftrightarrow\ & 0 \le \Big[-\nu_{i^*(k)} + \widehat{\nu}_{k,i^*(k)}(t-1) - S_{k,i^*(k)}(t-1)\Big] + \Big[\nu_{k,i} - \widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1)\Big] \\
&\quad + \Big[\nu_{i^*(k)} - \nu_{k,i} + 2S_{k,i}(t-1)\Big].
\end{aligned}$$

Thus, if we choose s_i such that the third term be non-positive, this will imply that one of the first two terms at least is non-negative.

Let s_i be defined by $s_i = \frac{20\log(T)}{\overline{\Delta}_{k,i}^2}$. Then, $O_{k,i}(t-1) \ge s_i$ implies $\nu_{i^*(k)} - \nu_{k,i} + 2S_{k,i}(t-1) \le 0$, that is the non-positivity of the third term. Thus, with this choice of s_i, if the inequality defining the first event in the probability holds, at least one of the first two terms above must be non-negative. In view of that, by the union bound and Hoeffding's inequality applied to the probability of each event, the following holds:

$$\begin{aligned}
r^2_{T,k,i} &\le \sum_{t=1}^{T} \mathbb{P}\left[0 \le -\nu_{i^*(k)} + \widehat{\nu}_{k,i^*(k)}(t-1) - S_{k,i^*(k)}(t-1) | \mathtt{A}_k\right] + \sum_{t=1}^{T} \mathbb{P}\left[0 \le \nu_{k,i} - \widehat{\nu}_{k,i}(t-1) - S_{k,i}(t-1)\Big|\mathtt{A}_k\right] \\
&\le 2\sum_{t=1}^{T} \frac{1}{t^4} \le 4.
\end{aligned}$$

Thus, this implies the inequality $\sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} r^2_{T,k,i} \le 4\big|\mathtt{A}_k \setminus \mathtt{B}_k\big|$. To bound $\sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} r^1_{T,k,i}$, we will use the clique covering $\mathcal{C}_k$ defined in Section 4. Since $\mathcal{C}_k$ is a cover of the graph G_k, we can decompose the expression involving $r^1_{T,k,i}$ in terms of the components of the clique cover and write

$$\sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}\, r^1_{T,k,i} \le \sum_{t=1}^{T} \sum_{C\in\mathcal{C}_k} \sum_{i\in C\setminus\mathtt{B}_k} \mathbb{E}[\overline{\Delta}_{k,i}\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}\mathbb{I}\{O_{k,i}(t-1) < s_i\}|\mathtt{A}_k].$$

Let $O_{k,C}(t-1)$ denote the number of times any expert in clique C has been played up to time $t-1$. Since experts in the same clique are observed together, $O_{k,C}(t-1)$ is less than or equal to the number of times an expert $i \in C$ is observed: $O_{k,C}(t-1) \le O_{k,i}(t-1)$. Thus, we can upper bound the expression above by replacing $O_{k,i}(t-1)$ with $O_{k,C}(t-1)$ as follows:

$$\begin{aligned}
\sum_{i\in\mathtt{A}_k\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}\, r^1_{T,k,i} &\le \sum_{t=1}^{T} \sum_{C\in\mathcal{C}_k} \sum_{i\in C\setminus\mathtt{B}_k} \mathbb{E}\left[\overline{\Delta}_{k,i}\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}\mathbb{I}\Big\{O_{k,C}(t-1) < s_i\Big\}\Big|\mathtt{A}_k\right] \\
&\le \sum_{t=1}^{T} \sum_{C\in\mathcal{C}_k} \sum_{i\in C\setminus\mathtt{B}_k} \mathbb{E}\left[\overline{\Delta}_{k,i}\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}\mathbb{I}\Big\{O_{k,C}(t-1) < \max_{i\in C\setminus\mathtt{B}_k} s_i\Big\}\Big|\mathtt{A}_k\right] \\
&\le \sum_{C\in\mathcal{C}_k} \Big(\max_{i\in C\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}\Big) \sum_{t=1}^{T} \sum_{i\in C\setminus\mathtt{B}_k} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\Big\{\nu_{k,i} > \nu_{i^*(k)}\Big\}\mathbb{I}\Big\{O_{k,C}(t-1) < \max_{i\in C\setminus\mathtt{B}_k} s_i\Big\}\Big|\mathtt{A}_k\right] \\
&\le \sum_{C\in\mathcal{C}_k} \Big(\max_{i\in C\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}\Big) \sum_{t=1}^{T} \sum_{i\in C\setminus\mathtt{B}_k} \mathbb{E}\left[\mathbb{I}\{I_t = i\}\mathbb{I}\Big\{O_{k,C}(t-1) < \max_{i\in C\setminus\mathtt{B}_k} s_i\Big\}\Big|\mathtt{A}_k\right] \\
&\le \sum_{C\in\mathcal{C}_k} \Big(\max_{i\in C\setminus\mathtt{B}_k} \overline{\Delta}_{k,i}\Big) \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{I_t \in C\}\mathbb{I}\Big\{O_{k,C}(t-1) < \max_{i\in C\setminus\mathtt{B}_k} s_i\Big\}\Big|\mathtt{A}_k\right].
\end{aligned}$$

Define s and t^* by $s = \max_{i\in C\setminus\mathtt{B}_k} s_i$ and $t^* = \max\Big\{t \le T\colon \mathbb{I}\{O_{k,C}(t-1) < s\} \ne 0\Big\}$. Then, we have

$$\sum_{t=1}^{T} \mathbb{I}\{I_t \in C\}\mathbb{I}\Big\{O_{k,C}(t-1) < s\Big\} = \sum_{t=1}^{t^*} \mathbb{I}\{I_t \in C\}\mathbb{I}\Big\{O_{k,C}(t-1) < s\Big\} \le s,$$

where the last inequality holds since, by definition of t^*, the number of non-zero terms in the last sum it at most s. Thus, we have $\sum_{t=1}^{T} \mathbb{E}[\mathbb{I}\{I_t \in C\}\mathbb{I}\{O_{k,C}(t-1) < s\}|\mathtt{A}_k] \leq s$ and

$$\sum_{i \in \mathtt{A}_k \setminus \mathtt{B}_k} \overline{\Delta}_{k,i}\, r^1_{T,k,i} \leq \sum_{C \in \mathcal{C}_k} \Big(\max_{i \in C \setminus \mathtt{B}_k} \overline{\Delta}_{k,i} \Big) \Big(\max_{i \in C \setminus \mathtt{B}_k} s_i \Big) = 20 \sum_{C \in \mathcal{C}_k} \frac{\max_{i \in C \setminus \mathtt{B}_k} \overline{\Delta}_{k,i}}{\min_{i \in C \setminus \mathtt{B}_k} \overline{\Delta}^2_{k,i}} \log(T),$$

for any clique covering $\mathcal{C}_k$. Combining this inequality with the one for $r^2_{T,k,i}$ gives:

$$\begin{aligned} \mathrm{R}^{\text{SLEEP}}_T(\text{UCB-SLG}) = \sum_{k=1}^{p} p_k \mathrm{R}_{T,k} &\leq \sum_{k=1}^{p} p_k \sum_{i \in \mathtt{A}_k \setminus \mathtt{B}_k} \overline{\Delta}_{k,i} (r^1_{T,k,i} + r^2_{T,k,i}) \\ &\leq \sum_{k=1}^{p} p_k \left(20 \sum_{C \in \mathcal{C}_k} \frac{\max_{i \in C \setminus \mathtt{B}_k} \overline{\Delta}_{k,i}}{\min_{i \in C \setminus \mathtt{B}_k} \overline{\Delta}^2_{k,i}} \log(T) + 4|\mathtt{A}_k \setminus \mathtt{B}_k| \right). \end{aligned}$$

Taking the minimum of the right-hand side over all possible clique covering $\mathcal{C}_k$ completes the proof. □

D. UCB-ABS

In this section, we prove Theorem 4.

Proof. [Theorem 4] To alleviate notation, throughout this proof we replace $L(\cdot, z_t)$ by $L_t(\cdot)$, $\mathbb{E}[\cdot|A^t = \mathtt{A}_k]$ by $\mathbb{E}[\cdot|\mathtt{A}_k]$, and $\nu_{k,i^*(k)}$ by $\nu_{i^*(k)}$. By the same reasoning as in the proof of Theorem 3, for each $k \in [p]$ the following holds:

$$\mathrm{R}_{T,k} \leq \sum_{t=1}^{T} \mathbb{E}[\mathbb{I}\{\nu_{k,I_t} > \nu_{i^*(k)}\}(L_t(\xi_{I_t}) - L_t(\xi_{i^*(k)}))|\mathtt{A}_k] \,, \tag{5}$$

where $\mathrm{R}_{T,k}$ is defined as in that proof.

Next, in order to bound (5), we split the rounds $t \in [T]$ into three cases that need to be dealt with separately:

1. $\nu_{i^*(k)} \neq c$ and round t is such that $I_t \neq 0$;
2. $\nu_{i^*(k)} = c$ and round t is such that $I_t \neq 0$;
3. Round t is such that $I_t = 0$.

In Case 1, the algorithm will pick an expert that is not $i^*(k)$ if there exists an expert $i \neq 0$ that satisfies $\widehat{\nu}_{k,i}(t-1) \leq \widehat{\nu}_{i^*(k)}(t-1)$. We will use a Follow-The-Leader type argument based on Lemma 1 of Caron et al. (2012). On the other hand, in Case 2, the algorithm will pick an expert that is not $i^*(k)$ if there exists an expert $i \neq 0$ that satisfies $\widehat{\nu}_{k,i}(t-1) - S_k(t-1) \leq c$. We will use a UCB-type argument. Finally, for Case 3, it must be that $c \leq \widehat{\nu}_{i^*(k)}(t-1) - S_k(t-1)$, and we will show that the overall contribution to the regret can be upper bounded by a constant, independent of time horizon T.

Case 1. Since $L_t(\xi_i) - L_t(\xi_{i^*(k)})$ and $\mathbb{I}\{I_t = i\}$ are conditionally independent given[3] $\mathtt{A}_k$, we can decompose the expectation in (5):

$$\sum_{t=1}^{T} \sum_{i \in [K] \setminus (\{0\} \cup \mathtt{B}_k)} \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\Big](\nu_{k,i} - \nu_{i^*(k)})$$

and focus on bounding the number of times each arm $i \in [K] \setminus (\{0\} \cup \mathtt{B}_k)$ was pulled. Similarly to the proof of Theorem 3, we introduce the conditions $O_k(t-1) > s_i$ and $O_k(t-1) < s_i$ for some s_i to be chosen later:

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\Big] \\
&= \sum_{t=1}^{T} \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{O_k(t-1) < s_i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\Big] \qquad (6) \\
&+ \sum_{t=1}^{T} \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{O_k(t-1) > s_i\}\mathbb{I}\{\nu_{k,i} > \nu_{i^*(k)}\}|\mathtt{A}_k\Big] \qquad (7)
\end{aligned}$$

and we bound (7) by a constant that is independent of T. If expert $I_t = i$ where $i \neq \{0\}$ is chosen, then it must be that $\widehat{\nu}_{k,i}(t-1) \leq \widehat{\nu}_{i^*(k)}(t-1)$. Hence,

$$(7) \leq \sum_{t=1}^{T} \mathbb{P}[\widehat{\nu}_{k,i}(t-1) \leq \widehat{\nu}_{i^*(k)}(t-1), i \neq \{0\}, O_k(t-1) > s_i, \nu_{k,i} > \nu_{i^*(k)}|\mathtt{A}_k\Big] \,.$$

We then use Lemma 1 of Caron et al. (2012) to show that the empirical mean of the chosen expert cannot be less than that of the best expert $i^*(k)$ too often. This gives us

$$(7) \leq \sum_{t=1}^{T} 2e^{-s_i(\nu_{k,i} - \nu_{i^*(k)})^2/2} \,.$$

[3] Recall that $\mathtt{A}_k$ is determined by x_t.

Note that this lemma applies here because the loss observations are i.i.d. given $A^t = \mathtt{A}_k$ and since $O_k(t-1) > s_i$, we saw at least s_i observations of the losses. We then choose $s_i = \frac{2\log(T\Delta_{k,i}^2)}{\Delta_{k,i}^2}$, so that $\sum_{t=1}^T 2e^{-s_i(\nu_{k,i}-\nu_{i^*(k)})^2/2} = O(1)$. Lastly, the bound on (6) follows by the same covering argument as in the proof of Theorem 3.

Case 2. By a similar reasoning as in Case 1, the regret is bounded as follows :

$$\sum_{t=1}^T \sum_{i\in[K]\setminus(\{0\}\cup\mathtt{B}_k)} \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{\nu_{k,i} > c\}|\mathtt{A}_k\Big](\nu_{k,i} - c)\,.$$

Again, for each expert $i \in [K]\setminus(\{0\}\cup\mathtt{B}_k)$, we split this sum according to the conditions $O_k(t-1) > s_i$ and $O_k(t-1) < s_i$ for some s_i to be chosen later. That is,

$$\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{\nu_{k,i} > c\}|\mathtt{A}_k\Big] \\
&= \sum_{t=1}^T \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{O_k(t-1) < s_i\}\mathbb{I}\{\nu_{k,i} > c\}|\mathtt{A}_k\Big] \qquad (8)\\
&+ \sum_{t=1}^T \mathbb{E}\Big[\mathbb{I}\{I_t = i\}\mathbb{I}\{i \neq \{0\}\}\mathbb{I}\{O_k(t-1) > s_i\}\mathbb{I}\{\nu_{k,i} > c\}|\mathtt{A}_k\Big]\,. \qquad (9)
\end{aligned}$$

To bound term (9), since $\nu_{i^*(k)} = c$, if $I_t = i$ was the chosen expert, then it must be that $\widehat{\nu}_{k,i}(t-1) - S_k(t-1) \le c$. Thus,

$$\begin{aligned}
(9) &\le \sum_{t=1}^T \mathbb{P}\left(\widehat{\nu}_{k,i}(t-1) - S_k(t-1) \le c, c < \nu_{k,i}, O_k(t-1) > s_i \middle| \mathtt{A}_k\right)\\
&= \sum_{t=1}^T \mathbb{P}\left(0 < -\widehat{\nu}_{k,i}(t-1) + S_k(t-1) + c + \nu_{k,i} - \nu_{k,i} + 2S_k(t-1) - 2S_k(t-1), c < \nu_{k,i}, O_k(t-1) > s_i \middle| \mathtt{A}_k\right)\\
&= \sum_{t=1}^T \mathbb{P}\left(0 < \nu_{k,i} - \widehat{\nu}_{k,i}(t-1) - S_k(t-1) + c - \nu_{k,i} + 2S_k(t-1), c < \nu_{k,i}, O_k(t-1) > s_i \middle| \mathtt{A}_k\right)\,,
\end{aligned}$$

where as in the proof of Theorem 3, we introduced the terms $\nu_{k,i}$ and $S_k(t-1)$. By choosing $s_i = \frac{20\log T}{(\nu_{k,i}-c)^2}$, then the condition $O_k(t-1) > s_i$ implies that $c - \nu_{k,i} + 2S_k(t-1) \le 0$. This in turn implies that $0 < \nu_{k,i} - \widehat{\nu}_{k,i}(t-1) - S_k(t-1)$, and we bound the probability of this latter event by using a union bound and Hoeffding's inequality:

$$\sum_{t=1}^T \mathbb{P}[0 < \nu_{k,i} - \widehat{\nu}_{k,i}(t-1) - S_k(t-1)] \le \sum_{t=1}^T \sum_{s=1}^t \frac{1}{t^5} \le \sum_{t=1}^T \frac{1}{t^4} \le 2\,.$$

Again, the bound on (8) follows directly by the covering argument in the proof of Theorem 3.

Case 3. Consider the rounds t where the chosen expert is the all-abstain expert, $(I_t = 0)$ (and where $I_t \notin \mathtt{B}_k$). By the same reasoning as in the previous two cases, the regret in this case can be bounded as follows:

$$(5) \le \sum_{t=1}^T \mathbb{E}[\mathbb{I}\{\nu_{k,I_t} > \nu_{i^*(k)}\}\mathbb{I}\{I_t = 0\}|\mathtt{A}_k](c - \nu_{i^*(k)})\,.$$

If the all-abstain expert is chosen at time t, then it must be that $c \le \widehat{\nu}_{i^*(k)}(t-1) - S_k(t-1)$. Hence,

$$\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\Big[\mathbb{I}\{\nu_{k,I_t} > \nu_{i^*(k)}\}\mathbb{I}\{I_t = 0\}|\mathtt{A}_k\Big]\\
&\qquad \le \sum_{t=1}^T \mathbb{P}\Big(c \le \widehat{\nu}_{k,i^*(k)}(t-1) - S_k(t-1), c > \nu_{i^*(k)}\Big|\mathtt{A}_k\Big)\,. \qquad (10)
\end{aligned}$$

By following a similar logic as in proof of Theorem 3, we then introduce $\nu_{i^*(k)}$ and use the fact that $c > \nu_{i^*(k)}$:

$$
\begin{aligned}
(10) &\leq \sum_{t=1}^{T} \mathbb{P}\Big(0 \leq -\nu_{i^*(k)} + \widehat{\nu}_{k,i^*(k)}(t-1) - S_k(t-1) + \nu_{i^*(k)} - c, c > \nu_{i^*(k)} \Big| \mathtt{A}_k\Big) \\
&\leq \sum_{t=1}^{T} \mathbb{P}\Big(0 \leq -\nu_{i^*(k)} + \widehat{\nu}_{k,i^*(k)}(t-1) - S_k(t-1) \Big| \mathtt{A}_k\Big) \\
&\leq \sum_{t=1}^{T} \sum_{s=1}^{t} \frac{1}{t^5} \leq \sum_{t=1}^{T} \frac{1}{t^4} \leq 2\,,
\end{aligned}
$$

where in the third-last inequality we used a union bound in conjunction with Hoeffding's inequality.

Combining the inequalities corresponding to three cases above completes the proof. □

E. Further Experimental Results

In this section, we present further experimental results testing different aspects of our problem. The first set of figures present the experimental results of all datasets using the same experimental setup as Cortes et al. (2018). Figure 4 and Figure 5 show the results for all abstention costs for two datasets `eye` and `HIGGS`. These results show that UCB-ABS outperform UCB, UCB-NT, and UCB-GT on most datasets, and approaches the performance of FS. Even though the experiments were carried out for all abstention costs, to simplify exposition, we show the results for the rest of the datasets for abstention costs in $\{0.05, 0.5, 0.95\}$ in Figure 6 and Figure 7. Figure 8, Figure 9, and Figure 10 show the fraction of abstained points for each algorithm for different abstention costs. As expected, all algorithms tend to abstain more often when the cost of abstention is smaller. Lastly, we increased the number of predictions functions from 100 to 200 hyperplanes and increased the number of abstention regions from 21 to 41. We find that the performance of all algorithms improves slightly on some datasets. Figure 11 shows the results of these new experiments for the same set of datasets and abstention costs as in the main part of the paper.

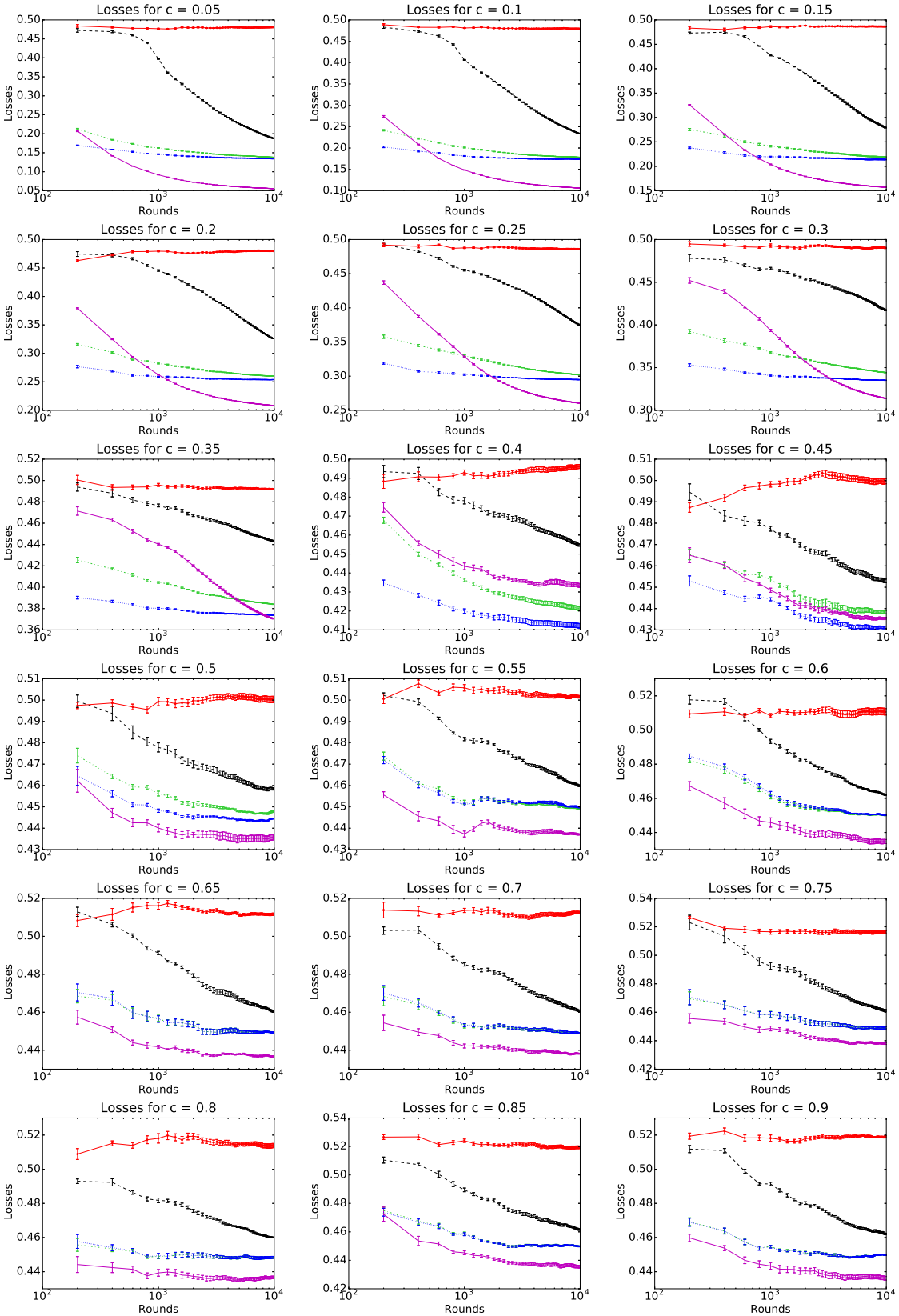

Figure 4: A graph of the averaged loss with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the `eye` dataset.

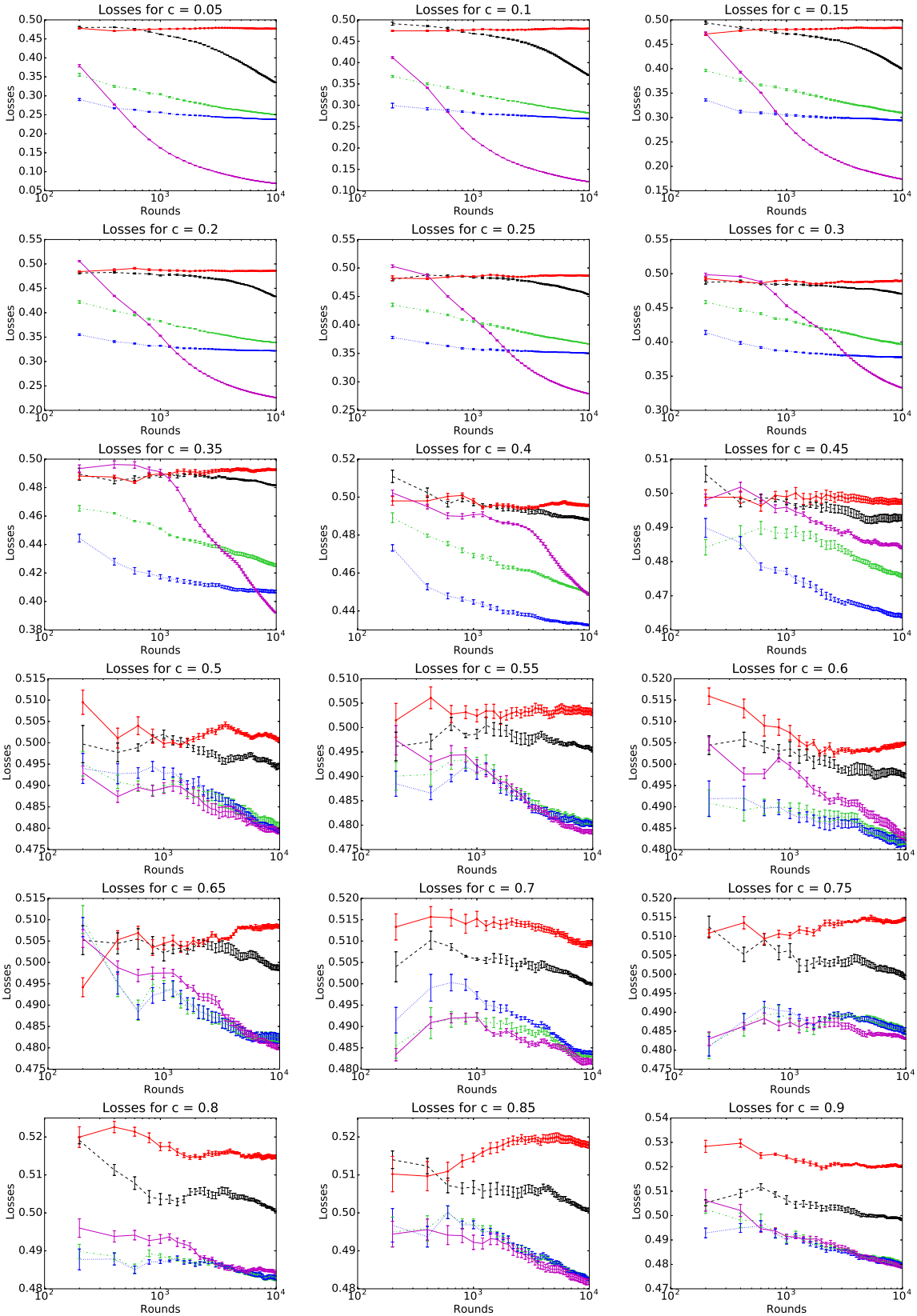

Figure 5: A graph of the averaged loss with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the HIGGS dataset.

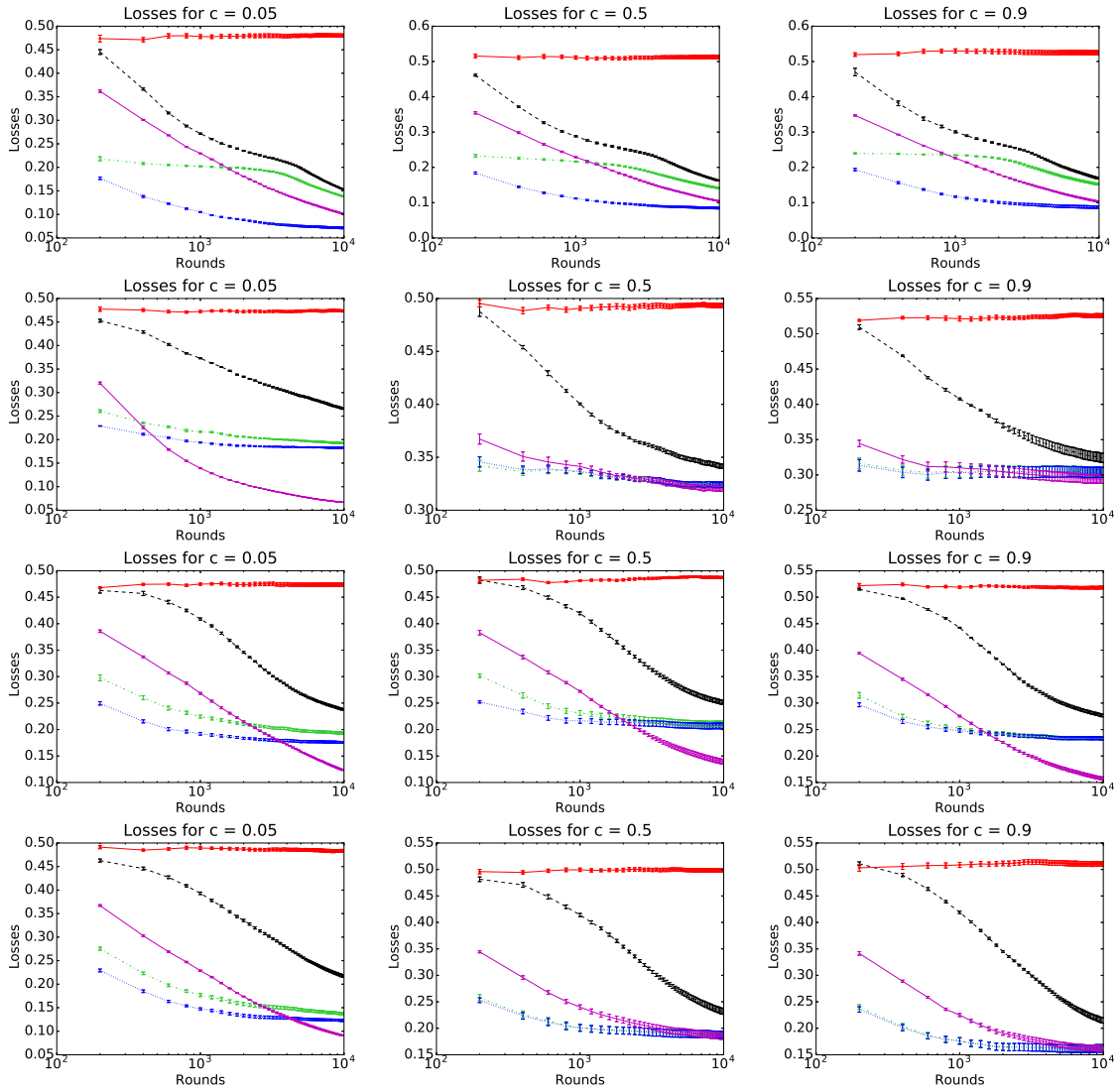

Figure 6: A graph of the averaged loss with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the `skin`, `cod-rna`, `guide`, `ijcnn` dataset.

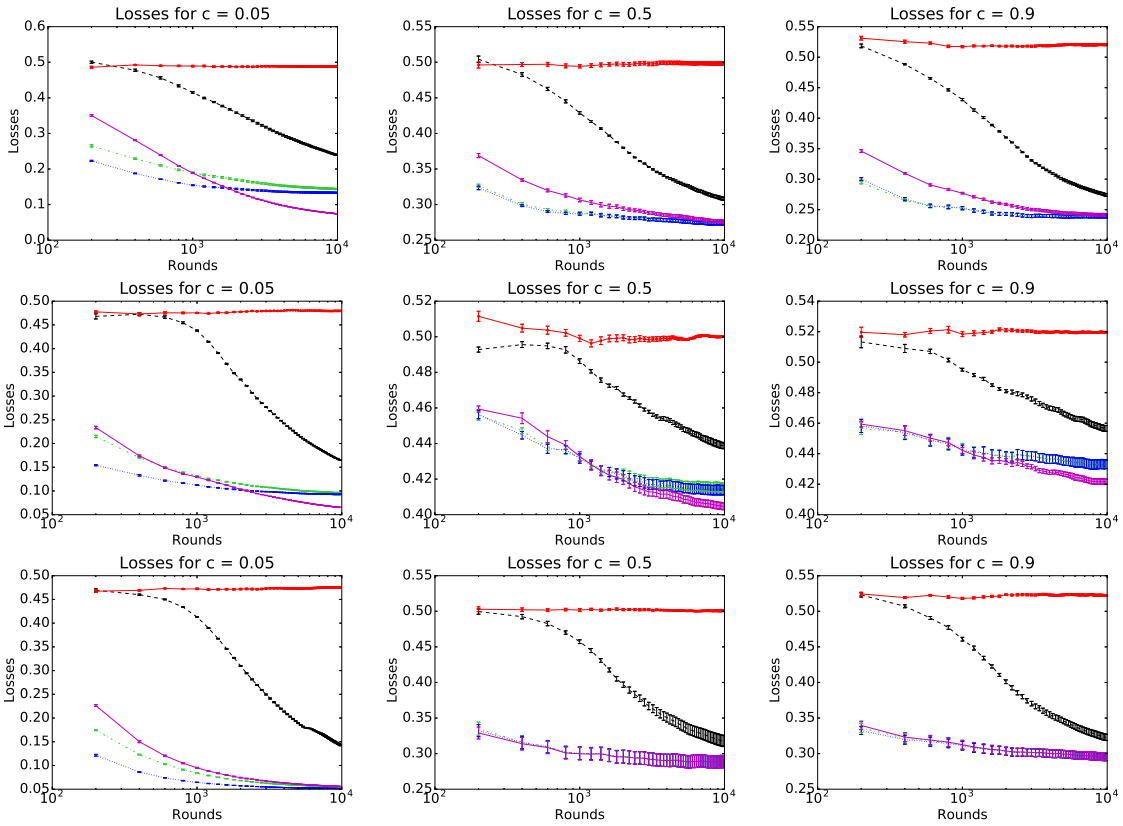

Figure 7: A graph of the averaged loss with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the CIFAR, covtype, and phish dataset.

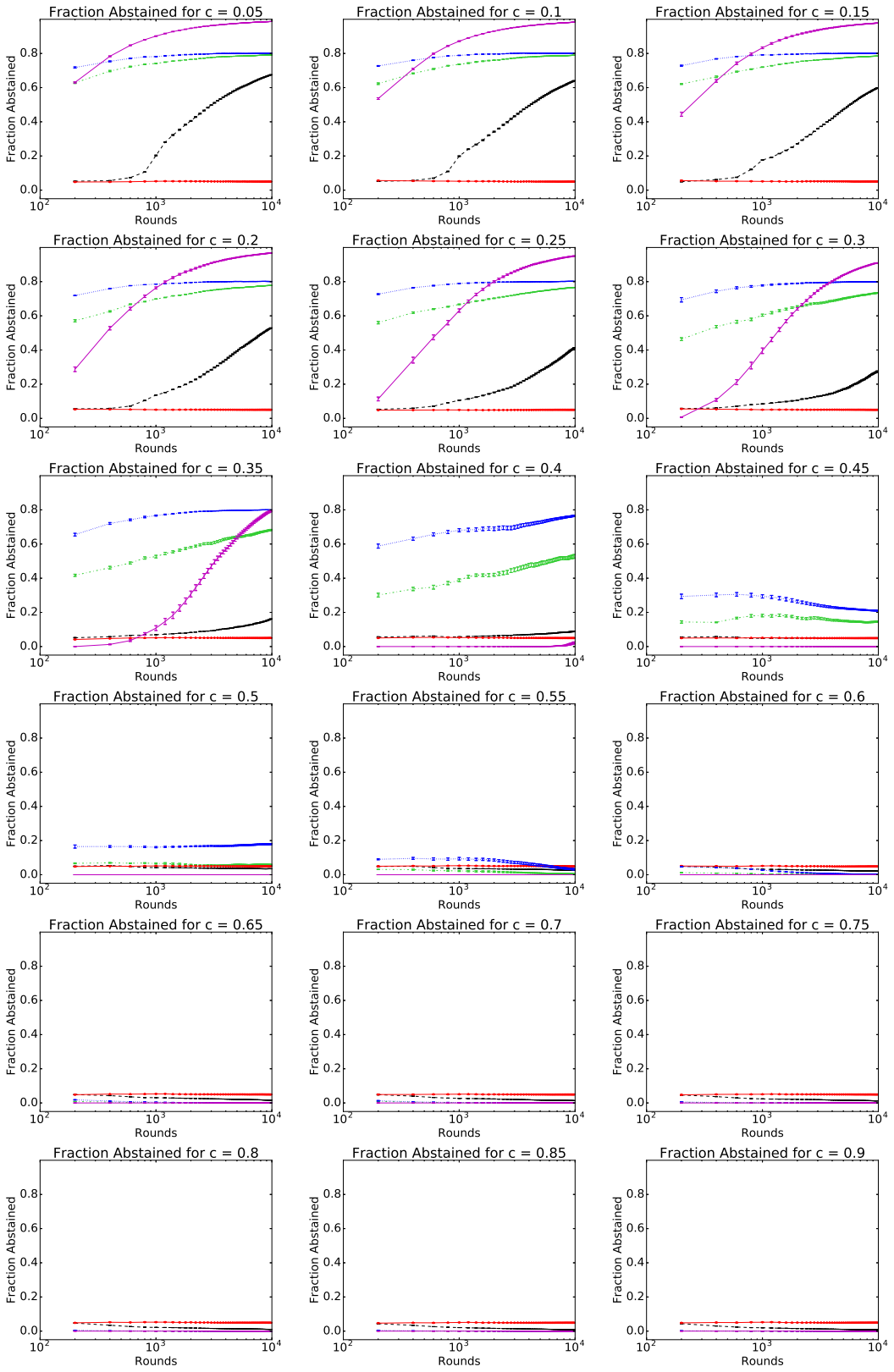

Figure 8: A graph of the fraction abstained with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the eye dataset.

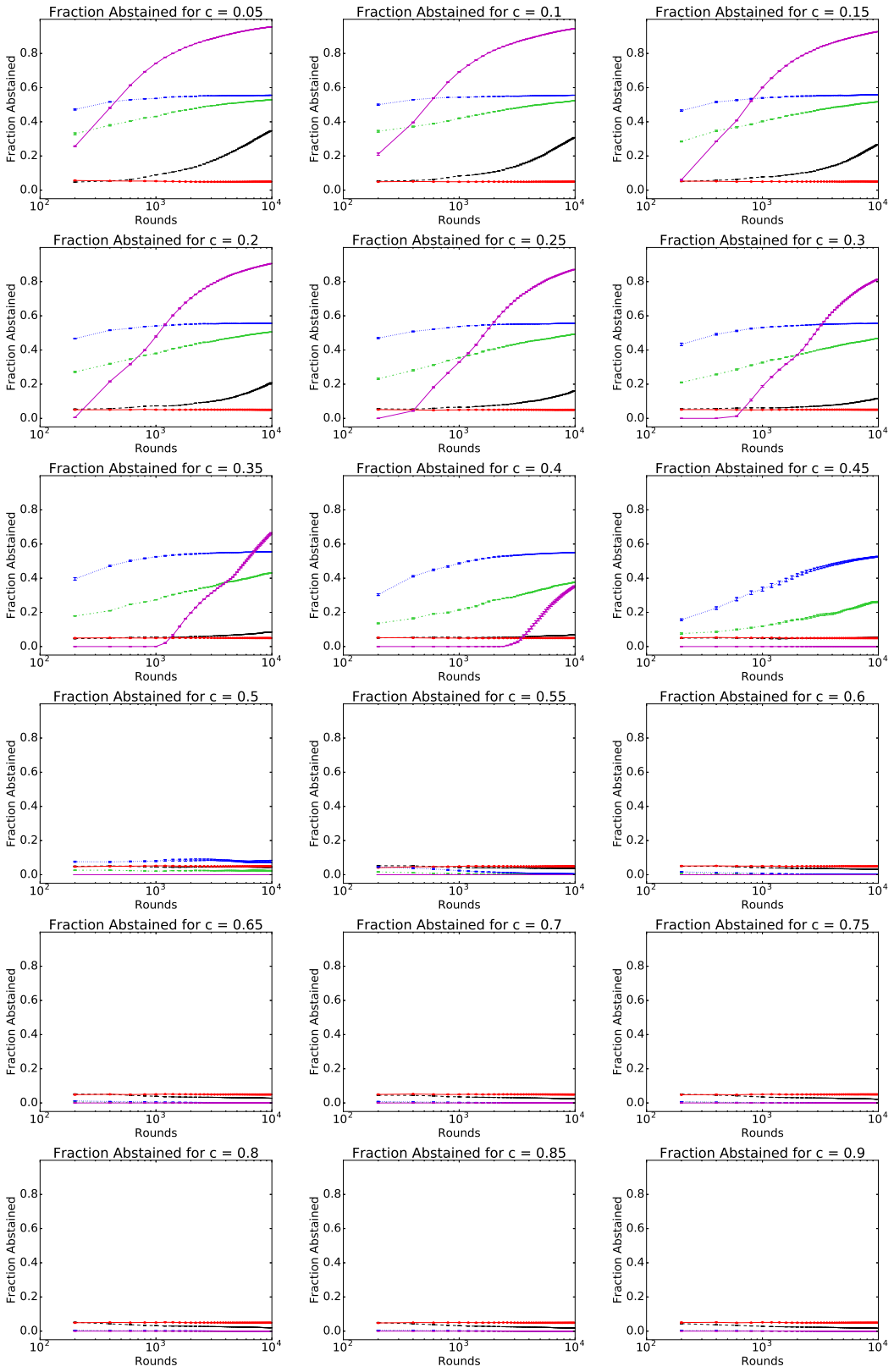

Figure 9: A graph of the fraction abstained with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. The results are for the HIGGS dataset.

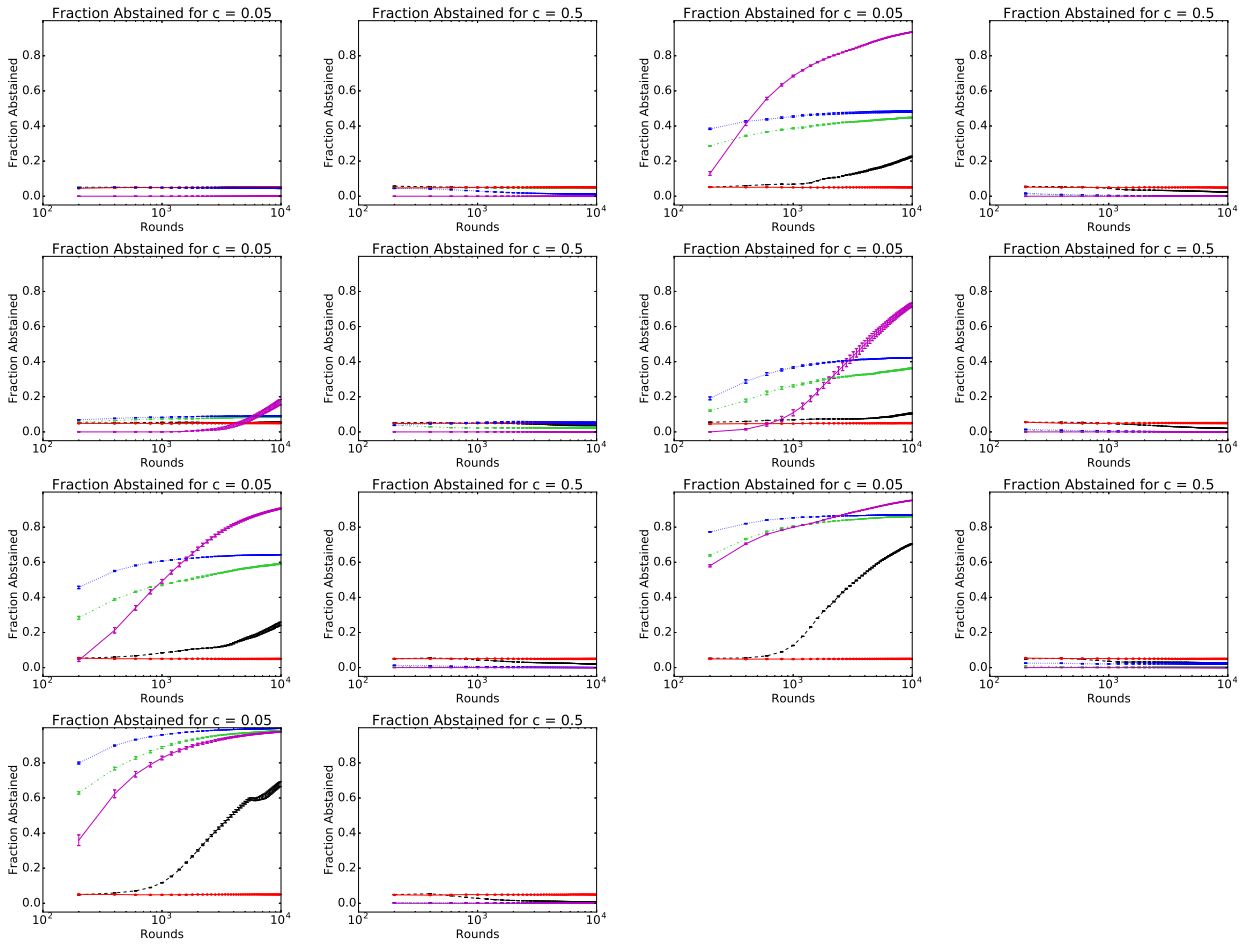

Figure 10: A graph of the fraction abstained with standard deviations as a function of t (log scale) for UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. We show the results for two abstention costs for each dataset. Starting from the top left, the plots are for the `skin` , `cod-rna`, `guide`, `ijcnn`, `CIFAR`, `covtype` and `phish` datasets.

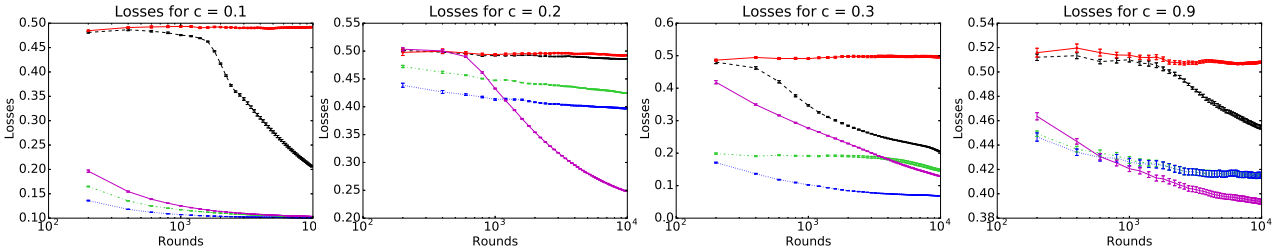

Figure 11: In this set of experiments, we increased the number of abstention functions from 21 to 41 and the number of hyperplanes from 100 to 200. The figures shows a graph of the averaged loss with standard deviations as a function of t (log scale). The algorithms we tested include UCB-ABS, UCB-GT, UCB-NT, UCB, and FS. Starting from the left, the datasets are as follows: `eye`, `HIGGS` , `skin`, and `covtype`.