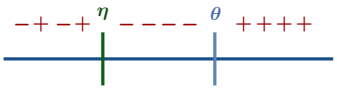

Figure 6: Simple example of the benefits of learning with abstention (Cortes et al., 2016a).

A. Further Related Work

Learning with abstention is a useful paradigm in applications where the cost of misclassifying a point is high. More concretely, suppose the cost of abstention c is less than $1/2$ and consider the set of points along the real line illustrated in Figure 6 where $+$ and $-$ indicate their labels. The best threshold classifier is the hypothesis given by threshold θ, since it correctly classifies points to the right of η, with an expected loss of $(1/2)\,\mathbb{P}[x \leqslant \eta]$. On the other hand, the best abstention pair (h, r) would abstain on the region left of η and correctly classify the rest, with an expected loss of $c\,\mathbb{P}(x \leqslant \eta)$. Since $c < 1/2$, the abstention pair always admits a better loss then the best threshold classifier.

Within the online learning literature, work related to our scenario includes the KWIK (*knows what it knows*) framework of Li et al. (2008) in which the learning algorithm is required to make only correct predictions but admits the option of abstaining from making a prediction. The objective is then to learn a concept exactly with the fewest number of abstentions. If in our framework we received the label at every round, KWIK could be seen as a special case of our framework for online learning with abstention with an infinite misclassification cost and some finite abstention cost. A relaxed version of the KWIK framework was introduced and analyzed by Sayedi et al. (2010) where a fixed number k of incorrect predictions are allowed with a learning algorithm related to the solution of the 'mega-egg game puzzle'. A theoretical analysis of learning in this framework was also recently given by Zhang & Chaudhuri (2016). Our framework does not strictly cover this relaxed framework. However, for some choices of the misclassification cost depending on the horizon, the framework is very close to ours. The analysis in these frameworks was given in terms of mistake bounds since the problem is assumed to be realizable. We will not restrict ourselves to realizable problems and, instead, will provide regret guarantees.

B. Additional material for the adversarial setting

We first present the pseudocode and proofs for the finite arm setting and next analyze the infinite arm setting.

B.1. Finite arm setting

Algorithm 3 contains the pseudocode for EXP3-ABS, an algorithm for online learning with abstention under an adversarial data model that guarantees small regret. The algorithm itself is a simple adaptation of the ideas in (Alon et al., 2014; 2015), where we incorporate the side information that the loss of an abstaining arm is always observed, while the loss of a predicting arm is observed only if the algorithm actually plays a predicting arm. In the pseudocode and in the proof that follows, $L_t(\xi_j)$ is a shorthand for $L(\xi_j, (x_t, y_t))$.

Proof of Theorem 1.
Proof. By applying the standard regret bound of Hedge (e.g., (Bubeck & Cesa-Bianchi, 2012)) to distributions $q_1, \ldots, q_T$ generated by EXP3-ABS and to the non-negative loss estimates $\widehat{L}_t(\xi_j)$, the following holds:

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\xi_j\in\mathcal{E}} q_t(\xi_j)\,\mathbb{E}\left[\widehat{L}_t(\xi_j)\right] - \sum_{t=1}^{T}\mathbb{E}\left[\widehat{L}_t(\xi^\star)\right]\right] \leqslant \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{\xi_j\in\mathcal{E}} q_t(\xi_j)\,\mathbb{E}\left[\widehat{L}_t(\xi_j)^2\right]\right], \tag{2}$$

for any fixed $\xi^\star \in \mathcal{E}$. Using the fact that $\mathbb{E}\left[\widehat{L}_t(\xi_j)\right] = L_t(\xi_j)$ and $\mathbb{E}\left[\widehat{L}_t(\xi_j)^2\right] = \frac{L_t(\xi_j)^2}{P_t(\xi_j)}$, we can write

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\xi_j\in\mathcal{E}} q_t(\xi_j)L_t(\xi_j) - \sum_{t=1}^{T}L_t(\xi^\star)\right] \leqslant \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{\xi_j\in\mathcal{E}} \frac{q_t(\xi_j)}{P_t(\xi_j)}L_t(\xi_j)^2\right].$$

For each t, we can split the nodes V of G_t^{ABS} into the two subsets $V_{abs,t}$ and $V_{acc,t}$ where if a node ξ_j is abstaining at time t

ALGORITHM 3: EXP3-ABS

input Set of experts $\mathcal{E} = \{\xi_1, \dots, \xi_K\}$; learning rate $\eta > 0$;
Init: q_1 is the uniform distribution over $\mathcal{E}$;
for $t \leftarrow 1, 2, \dots$ **do**
RECEIVE(x_t);
$\xi_{I_t} \leftarrow$ SAMPLE(q_t);
if $r_{I_t}(x_t) > 0$ **then**
RECEIVE(y_t);
end if
For all $\xi_j = (h_j, r_j)$, set :

$$P_t(\xi_j) \leftarrow \begin{cases} 1 & \text{if } r_j(x_t) \leqslant 0 \\ \sum_{\xi_i \in \mathcal{E} \,:\, r_i(x_t) > 0} q_t(\xi_i) & \text{if } r_j(x_t) > 0 \,, \end{cases}$$

$$\widehat{L}_t(\xi_j) \leftarrow \frac{L_t(\xi_j)}{P_t(\xi_j)} \left(1_{r_{I_t}(x_t) \leqslant 0} 1_{r_j(x_t) \leqslant 0} + 1_{r_{I_t}(x_t) > 0} \right) \,,$$

$$q_{t+1}(\xi_j) \leftarrow \frac{q_t(\xi_j) \exp(-\eta \widehat{L}_t(\xi_j))}{\sum_{\xi_i \in \mathcal{E}} q_t(\xi_i) \exp(-\eta \widehat{L}_t(\xi_i))} \,.$$

end for

ALGORITHM 4: CONTEXP3-ABS.

input Ball radius $\varepsilon > 0$, ε-covering $\mathcal{Y}_\varepsilon$ of $\mathcal{Y}$ such that $|\mathcal{Y}_\varepsilon| \leq C_\mathcal{Y}\, \varepsilon^{-2}$;
for $t = 1, 2, \dots$ **do**
RECEIVE(x_t);
If x_t does not belong to any existing ball, create new ball of radius ε centered on x_t, and allocate fresh instance of EXP3-ABS;
Let "Active EXP3-ABS" be the instance allocated to the existing ball whose center x_s is closest to x_t;
Draw action $\xi_{I_t} \in \mathcal{Y}_\varepsilon$ using Active EXP3-ABS;
Get loss feedback associated with ξ_{I_t} and use it to update state of "Active EXP3-ABS".
end for

then $\xi_j \in V_{abs,t}$, and otherwise $\xi_j \in V_{acc,t}$. Thus, for any round t, we can write

$$\begin{aligned} \sum_{\xi_j \in \mathcal{E}} \frac{q_t(\xi_j)}{P_t(\xi_j)} L_t(\xi_j)^2 &= \sum_{\xi_j \in V_{abs,t}} \frac{q_t(\xi_j)}{P_t(\xi_j)} L_t(\xi_j)^2 + \sum_{\xi_j \in V_{acc,t}} \frac{q_t(\xi_j)}{P_t(\xi_j)} L_t(\xi_j)^2 \\ &\leqslant \sum_{\xi_j \in V_{abs,t}} q_t(\xi_j)\, c^2 + \sum_{\xi_j \in V_{acc,t}} \frac{q_t(\xi_j)}{P_t(\xi_j)} \\ &\leqslant c^2 + 1 \,. \end{aligned}$$

The first inequality holds since if ξ_j is an abstaining expert at time t, we know that $L_t(\xi_j) = c$ and $P_t(\xi_j) = 1$, while for the accepting experts we know that $L_t(\xi_j) \leqslant 1$ anyway. The second inequality holds because if ξ_j is an accepting expert, we have $P_t(\xi_j) = \sum_{\xi_j \in V_{acc,t}} q_t(\xi_j)$. Combining this inequality with (2) concludes the proof. □

B.2. Infinite arm setting

Here, the input space $\mathcal{X}$ is assumed to be totally bounded, so that there exists a constant $C_\mathcal{X} > 0$ such that, for all $0 < \varepsilon \leqslant 1$, $\mathcal{X}$ can be covered with at most $C_\mathcal{X} \varepsilon^{-d}$ balls of radius ε. Let $\mathcal{Y}$ be a shorthand for $[-1, 1]^2$, the range space of the pairs (h, r). An ε-covering $\mathcal{Y}_\varepsilon$ of $\mathcal{Y}$ with respect to the Euclidean distance on $\mathcal{Y}$ has size $K_\varepsilon \leqslant C_\mathcal{Y} \varepsilon^{-2}$ for some constant $C_\mathcal{Y}$.

The online learning scenario for the loss $\widetilde{L}$ under the abstention setting's feedback graphs is as follows. Given an unknown sequence $z_1, z_2, \dots$ of pairs $z_t = (x_t, y_t) \in \mathcal{X} \times \{\pm 1\}$, for every round $t = 1, 2, \dots$:

1. The environment reveals input $x_t \in \mathcal{X}$;
2. The learner selects an action $\xi_{I_t} \in \mathcal{Y}$ and incurs loss $\widetilde{L}(\xi_{I_t}, z_t)$;
3. The learner obtains feedback from the environment.

Our algorithm is described as Algorithm 4. The algorithm essentially works as follows. At each round t, if a new incoming input $x_t \in \mathcal{X}$ is not contained in any existing ball generated so far, then a new ball centered at x_t is created, and a new instance of EXP3-ABS is allocated to handle x_t. Otherwise, the EXP3-ABS instance associated with the closest input so far is used. Each allocated EXP3-ABS instance operates on the discretized action space $\mathcal{Y}_\varepsilon$.

Consider the function

$$\widetilde{L}(a,r) = \begin{cases} c & \text{if } r \leqslant -\gamma \\ 1 + \left(\frac{1-c}{\gamma}\right) r & \text{if } r \in (-\gamma, 0) \\ 1 - \left(\frac{1-f_\gamma(-a)}{\gamma}\right) r & \text{if } r \in [0, \gamma) \\ f_\gamma(-a) & \text{if } r \geqslant \gamma\,, \end{cases}$$

where f_γ is the Lipschitz variant of the 0/1-loss mentioned in Section 3 of the main text (Figure 2 (a)). For any fixed a, the function $\widetilde{L}(a,r)$ is $1/\gamma$-Lipschitz when viewed as a function of r, and is $1/(2\gamma)$-Lipschitz for any fixed r when viewed as a function of a. Hence

$$\begin{aligned} |\widetilde{L}(a,r) - \widetilde{L}(a',r')| &\leqslant |\widetilde{L}(a,r) - \widetilde{L}(a,r')| + |\widetilde{L}(a,r') - \widetilde{L}(a',r')| \\ &\leqslant \frac{1}{\gamma}\,|r - r'| + \frac{1}{2\gamma}\,|a - a'| \\ &\leqslant \sqrt{\frac{1}{\gamma^2} + \frac{1}{4\gamma^2}}\,\sqrt{(a-a')^2 + (r-r')^2} \\ &< \frac{2}{\gamma}\,\sqrt{(a-a')^2 + (r-r')^2}\,, \end{aligned}$$

so that $\widetilde{L}$ is $\frac{2}{\gamma}$-Lipschitz w.r.t. the Euclidean distance on $\mathcal{Y}$. Furthermore, a quick comparison to the abstention loss

$$L(a,r) = f_\gamma(a) 1_{r>0} + c 1_{r\leqslant 0}$$

reveals that (recall Figure 2 (b) in the main text) :

- $\widetilde{L}$ is an upper bound on L, i.e.,
$$\widetilde{L}(a,r) \geqslant L(a,r), \quad \forall\,(a,r) \in \mathcal{Y}\,;$$
- $\widetilde{L}$ approximates L in that
$$\widetilde{L}(a,r) = L(a,r), \quad \forall\,(a,r) \in \mathcal{Y} \,:\, |r| \geqslant \gamma\,. \tag{3}$$

With the above properties of $\widetilde{L}$ at hand, we are ready to prove Theorem 2.

Proof of Theorem 2.
Proof. On each ball $B \subseteq \mathcal{X}$ that CONTEXP3-ABS allocates during its online execution, Theorem 1 supplies the following regret guarantee for the associated instance of EXP3-ABS:

$$\frac{\log K_\varepsilon}{\eta} + \frac{\eta}{2}\,T_B(c^2+1)\,,$$

where T_B is the number of points x_t falling into ball B. Now, taking into account that $\widetilde{L}$ is $\frac{2}{\gamma}$-Lipschitz, and that the functions h and r are assumed to be $L_\mathcal{E}$-Lipschitz on $\mathcal{X}$, a direct adaptation of the proof of Theorem 1 in (Cesa-Bianchi et al., 2017) gives the bound

$$\sup_{\xi\in\mathcal{E}} \mathbb{E}\left[\sum_{t=1}^T \widetilde{L}(\xi_{I_t}, z_t) - \sum_{t=1}^T \widetilde{L}(\xi, z_t)\right] \leqslant \frac{N_T \log K_\varepsilon}{\eta} + \frac{\eta}{2}\,T(c^2+1) + L_\mathcal{E}\,\varepsilon\,\frac{2}{\gamma}\,T\,,$$

being $N_T \leqslant C_{\mathcal{X}} \varepsilon^{-d}$ the maximum number of balls created by CONTEXP3-ABS. Using $c \leqslant 1$ and setting $\eta = \sqrt{\frac{N_T \log K_\varepsilon}{T}}$ yields

$$\sup_{\xi \in \mathcal{E}} \mathbb{E}\left[\sum_{t=1}^{T} \widetilde{L}(\xi_{I_t}, z_t) - \sum_{t=1}^{T} \widetilde{L}(\xi, z_t)\right] \leqslant 2\sqrt{T\, N_T\, \log K_\varepsilon} + L_{\mathcal{E}}\, \varepsilon\, \frac{2}{\gamma}\, T\ .$$

Next, optimizing for ε by setting $\varepsilon \simeq T^{-\frac{1}{2+d}} \left(\frac{1}{\gamma}\right)^{-\frac{2}{2+d}}$ (and disregarding $L_{\mathcal{E}}$ and log factors) gives

$$\sup_{\xi \in \mathcal{E}} \mathbb{E}\left[\sum_{t=1}^{T} \widetilde{L}(\xi_{I_t}, z_t) - \sum_{t=1}^{T} \widetilde{L}(\xi, z_t)\right] = \widetilde{\mathcal{O}}\left(T^{\frac{d+1}{d+2}} \left(\frac{1}{\gamma}\right)^{\frac{d}{d+2}}\right)\ . \tag{4}$$

Finally, we are left with connecting the above bound on the regret with a bound on the regret for L. Now, observe that

$$\mathbb{E}\left[\sum_{t=1}^{T} \widetilde{L}(\xi_{I_t}, z_t)\right] \geqslant \mathbb{E}\left[\sum_{t=1}^{T} L(\xi_{I_t}, z_t)\right]\ , \tag{5}$$

since $\widetilde{L}(\xi, z_t)$ is an upper bound on $L(\xi, z_t)$ for any ξ and z_t. Moreover, if we assume for the sake of brevity that the minima are reached (the general case is straightforward to handle in a similar way), we can define

$$\xi^* = (h^*, r^*) = \operatorname*{argmin}_{\xi \in \mathcal{E}} \sum_{t=1}^{T} L(\xi, z_t), \qquad \widetilde{\xi}^* = \operatorname*{argmin}_{\xi \in \mathcal{E}} \sum_{t=1}^{T} \widetilde{L}(\xi, z_t)\ .$$

We denote by $M_T^*(\gamma)$ the number of x_t such that $|r^*(x_t)| \leqslant \gamma$. Then, we can write

$$\begin{aligned}
\sum_{t=1}^{T} \widetilde{L}(\widetilde{\xi}^*, z_t) &\leqslant \sum_{t=1}^{T} \widetilde{L}(\xi^*, z_t) \\
&\leqslant \sum_{t\,:\,|r^*(x_t)|>\gamma}^{T} \widetilde{L}(\xi^*, z_t) + M_T^*(\gamma) \\
&\quad \text{(since } \widetilde{L} \leqslant 1\text{)} \\
&= \sum_{t\,:\,|r^*(x_t)|>\gamma}^{T} L(\xi^*, z_t) + M_T^*(\gamma) \\
&\quad \text{(using (3))} \\
&\leqslant \sum_{t=1}^{T} L(\xi^*, z_t) + M_T^*(\gamma)\ .
\end{aligned}$$

Combining with (4) and (5) gives the following regret bound

$$\sup_{\xi \in \mathcal{E}} \mathbb{E}\left[\sum_{t=1}^{T} L(\xi_{I_t}, z_t) - \sum_{t=1}^{T} L(\xi, z_t)\right] \leqslant \widetilde{\mathcal{O}}\left(T^{\frac{d+1}{d+2}} \left(\frac{1}{\gamma}\right)^{\frac{d}{d+2}}\right) + M_T^*(\gamma)\ ,$$

thereby concluding the proof. □

Remark 1 *The reader should observe that, since the algorithm is competing against an* uncountably infinite *set of experts, the standard regret guarantee of $\sqrt{T}$ that one can achieve in the finite case cannot be obtained in general (see, e.g., the lower bound on regret of $T^{(d-1)/d}$ by (Hazan & Megiddo, 2007), which holds in the easier full information setting). Notice that, while our algorithm* CONTEXP3-ABS *admits a slightly worse bound of the form $T^{(d+1)/(d+2)}$, it has the advantage of being computationally feasible. In particular, the covering of the input space $\mathcal{X}$ can be done adaptively, as the points x_t are observed. In doing so, the number of ε-balls allocated can never exceed the total number of rounds T. Given a new x_t, the algorithm has to decide if a new ball needs to be created or an old ball can be used. Known data-structures exist to efficiently implement this decision (e.g., (Clarkson, 2006)). The extra additive term $M_T^*(\gamma)$ in Theorem 2 is due to the*

fact that the loss function L therein is not Lipschitz. In fact, one can further improve the term $T^{\frac{d+1}{d+2}}$ to $T^{\frac{d}{d+1}}$ by adopting a hierarchical covering technique of the function space $\mathcal{E}$, each layer of the hierarchy being a pool of experts for the layer above it, see, e.g., (Cesa-Bianchi et al., 2017). However, the resulting algorithm would be of theoretical interest only, since it would be computationally very costly.

C. Additional material for the stochastic setting

In this section, we present the proofs of the theoretical guarantees for UCB-NT and UCB-GT, as well as the proof of Proposition 1. The following theorems hold more generally with $S_{j,t} = \sqrt{\frac{2\beta \log t}{Q_{j,t}}}$ for $\beta > 2$, which implies slightly better constants in the regret bound. However, for the sake of the simplicity of the presentation, below we set $\beta = \frac{5}{2}$. Moreover, we prove Theorem 3 for the abstention loss L, but it holds for any general loss function.

C.1. Regret of UCB-NT

We now prove the theorem for UCB-NT based on the admissible p-partitioning of the time-varying feedback graphs.

Proof of Theorem 3.
Proof. Consider a sequence of graph realizations $G_1, \ldots, G_t$ denoted by $\mathbf{G}_t$. By conditioning on this quantity, the regret can be decomposed according to each arm i:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[L(\xi_{I_t}, z_t) - L(\xi_*, z_t)] &= \sum_{t=1}^{T} \mathbb{E}[\mathbb{E}[L(\xi_{I_t}, z_t) - L(\xi_*, z_t)|\mathbf{G}_t]] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{K} 1_{I_t=i}(L(\xi_i, z_t) - L(\xi_*, z_t))\middle|\mathbf{G}_t\right]\right] \\
&= \sum_{i=1}^{K}\sum_{t=1}^{T} \mathbb{E}[\mathbb{E}[L(\xi_i, z_t) - L(\xi_*, z_t)|\mathbf{G}_t]\,\mathbb{E}[1_{I_t=i}|\mathbf{G}_t]] \\
&= \sum_{i=1}^{K}\sum_{t=1}^{T} \mathbb{E}[\mathbb{E}[L(\xi_i, z_t) - L(\xi_*, z_t)]\,\mathbb{E}[1_{I_t=i}|\mathbf{G}_t]] = \mathbb{E}\left[\sum_{i=1}^{K}\sum_{t=1}^{T} \Delta_i\,\mathbb{E}[1_{I_t=i}|\mathbf{G}_t]\right]
\end{aligned}$$

where, in the last step, we used the fact that $L(\cdot, z_t)$s are independent of $\mathbf{G}_t$ since, by assumption, $\mathbf{G}_t$ only depends on information up $t-1$. Next, we focus on bounding $\sum_{t=1}^{T} \mathbb{E}[1_{I_t=i}|\mathbf{G}_t]$ for each arm i.

We split the expectation according to the events $Q_{i,t-1} > s_i$ and $Q_{i,t-1} \leqslant s_i$, where s_i is a quantity determined later:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[1_{I_t=i}|\mathbf{G}_t] &= \sum_{t=1}^{T} \mathbb{E}[1_{I_t=i}(1_{Q_{i,t-1}\leqslant s_i} + 1_{Q_{i,t-1}>s_i})|\mathbf{G}_t] \\
&\leqslant s_i + \sum_{t=1}^{T} \mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}>s_i}|\mathbf{G}_t].
\end{aligned}$$

We wish to choose s_i sufficiently large so that the second term is bounded but so that it admits a mild dependence on T. Now, whenever $I_t = i$, by the design of the algorithm, it must be the case that the upper confidence bound of i is smaller than that of any other expert. Thus,

$$\mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}>s_i}|\mathbf{G}_t] = \mathbb{P}[I_t = i, Q_{i,t-1} > s_i|\mathbf{G}_t] \leqslant \mathbb{P}[\widehat{\mu}_{i,t-1} - S_{i,t-1} \leqslant \widehat{\mu}_{*,t-1} - S_{*,t-1}, Q_{i,t-1} > s_i|\mathbf{G}_t],$$

where $*$ denotes the best-in-class expert. We now use the terms μ_*, μ_i and $S_{i,t-1}$ to reorder the first event in the probability on the right-hand side of the last expression as follows:

$$\begin{aligned}
&0 \leqslant \widehat{\mu}_{*,t-1} - S_{*,t-1} - \widehat{\mu}_{i,t-1} + S_{i,t-1} \\
\Leftrightarrow\ &0 \leqslant (\widehat{\mu}_{*,t-1} - S_{*,t-1} - \mu_*) + (\mu_i - \widehat{\mu}_{i,t-1} + S_{i,t-1} - 2S_{i,t-1}) + (\mu_* - \mu_i + 2S_{i,t-1})\,.
\end{aligned}$$

If we can show that the third term is negative, then the first and second term must be positive. Moreover, we will further show that the first and second terms can only be positive with an extremely low probability that is bounded by a constant

independent of T. Furthermore, the third term will be negative whenever the slack term in the upper confidence bound is small enough, which amounts to choosing s_i large enough.

In particular, by setting $s_i = \frac{20\log(T)}{\Delta_i^2}$, we ensure that the event $Q_{i,t-1} > s_i$ implies that

$$Q_{i,t-1} > \frac{20\log(t)}{\Delta_i^2} \Leftrightarrow \mu_* - \mu_i + 2S_{i,t-1} < 0.$$

As explained above, it then follows that

$$\begin{aligned}
&\mathbb{P}[\widehat{\mu}_{i,t-1} - S_{i,t-1} \leqslant \widehat{\mu}_{*,t-1} - S_{*,t-1}, Q_{i,t-1} > s_i | \mathbf{G}_t] \\
&\leqslant \mathbb{P}[\widehat{\mu}_{*,t-1} - S_{*,t-1} - \mu_* \geqslant 0 | \mathbf{G}_t] + \mathbb{P}[\mu_i - \widehat{\mu}_{i,t-1} + S_{i,t-1} - 2S_{i,t-1} \geqslant 0 | \mathbf{G}_t].
\end{aligned}$$

We can bound these last probabilities using the union bound and a concentration inequality such as Hoeffding's Inequality:

$$\begin{aligned}
&\mathbb{P}[\mu_i - \widehat{\mu}_{i,t-1} + S_{i,t-1} - 2S_{i,t-1} \geqslant 0 | \mathbf{G}_t] \\
&= \mathbb{P}\Big[- \tfrac{1}{Q_{i,t-1}} \sum_{s=1}^{t-1} L(\xi_i, z_s) 1_{i \in N_s(I_s)} + \mu_i - \sqrt{\tfrac{5\log(t)}{Q_{i,t-1}}} \geqslant 0 \Big| \mathbf{G}_t \Big].
\end{aligned}$$

Now, the estimate $\widehat{\mu}_{i,t-1}$ is an average of i.i.d. realizations of the random variable $L(\xi_i, z)$, with $z \sim \mathcal{D}$, since the out-neighborhood of the chosen expert only depends on previous observations. That is,

$$\begin{aligned}
\frac{\mathbb{E}[\sum_{s=1}^{t-1} L(\xi_i, z_s) 1_{i \in N_s(I_s)}]}{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)}]} &= \frac{\mathbb{E}[\sum_{s=1}^{t-1} \mathbb{E}[L(\xi_i, z_s) 1_{i \in N_s(I_s)} | i \in N_s(I_s)]]}{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)}]} \\
&= \frac{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)} \mathbb{E}[L(\xi_i, z_s) | i \in N_s(I_s)]]}{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)}]} \\
&= \frac{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)} \mathbb{E}[L(\xi_i, z_s)]]}{\mathbb{E}[\sum_{s=1}^{t-1} 1_{i \in N_s(I_s)}]} \\
&= \mathbb{E}[L(\xi_i, z)].
\end{aligned}$$

Hence, $\widehat{\mu}_{i,t-1}$ can be turned into an empirical estimate of μ_i using the union bound as follows:

$$\begin{aligned}
\mathbb{P}\Big[- \tfrac{1}{Q_{i,t-1}} \sum_{s=1}^{t-1} L(\xi_i, z_s) 1_{i \in N_s(I_s)} + \mu_i - \sqrt{\tfrac{5\log(t)}{Q_{i,t-1}}} \geqslant 0 \Big| \mathbf{G}_t \Big] &\leqslant \mathbb{P}\Big[\exists n \in [1,t] : -\widehat{\mu}_i^n + \mu_i - \sqrt{\tfrac{5\log(t)}{n}} \Big| \mathbf{G}_t \Big] \\
&\leqslant \sum_{n=1}^{t} \frac{1}{t^{\frac{5}{2}}} = \frac{1}{t^{\frac{3}{2}}},
\end{aligned}$$

where $\widehat{\mu}_i^n = \frac{1}{n} \sum_{s=1}^{n} L(\xi_i, z_s)$. By the same reasoning, we can also bound the probability of the best arm :

$$\mathbb{P}\Big[\widehat{\mu}_{*,t-1} - S_{*,t-1} - \mu_* \geqslant 0 \Big| \mathbf{G}_t \Big] \leqslant \sum_{n=1}^{t} \frac{1}{t^{\frac{5}{2}}} = \frac{1}{t^{\frac{3}{2}}}.$$

By assumption, for each $\mathcal{C}_{t,k}$, $\forall i, j \in \mathcal{C}_{t,k}$, it is the case that $i \in N_t(j)$. Moreover, for any $i \in [K]$, it must be the case that for every $s \in [t]$, $i \in \mathcal{C}_{s,k}$ for some $k \in [p]$. With these clusters $\mathcal{C}_{s,k}$, we can write

$$Q_{i,t} = \sum_{s=1}^{t} 1_{i \in N_s(I_s)} = \sum_{s=1}^{t} \sum_{j=1}^{K} 1_{i \in N_s(j)} 1_{I_s = j} \geqslant \sum_{s=1}^{t} \sum_{j \in \mathcal{C}_{s,k}} 1_{i \in N_s(j)} 1_{I_s = j} = \sum_{s=1}^{t} \sum_{j \in \mathcal{C}_{s,k}} 1_{I_s = j},$$

which implies that

$$\sum_{i=1}^{K}\sum_{t=1}^{T}\Delta_i\,\mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}\leqslant s_i}|\mathbf{G}_t] \leqslant \sum_{k\in[p]}\left[\max_{\substack{t\in[T]\\ j\in\mathcal{C}_{t,k}}}\Delta_j\right]\sum_{t=1}^{T}\sum_{j\in\mathcal{C}_{t,k}}\mathbb{E}\left[1_{I_t=j}1_{Q_{j,t-1}\leqslant\max_{\substack{t\in[T]\\ j\in\mathcal{C}_{t,k}}} s_j}\middle|\mathbf{G}_t\right]$$
$$\leqslant \sum_{k\in[p]}\left[\max_{t\in[T]}\max_{j\in\mathcal{C}_{t,k}}\Delta_j\right]\left[\max_{t\in[T]}\max_{j\in\mathcal{C}_{t,k}} s_j\right].$$

Combining the above calculations, applying our definition for s_i, and using the fact that the above analysis holds for any such partition shows that

$$\mathbb{E}\left[\min_{\substack{p\\ \{\mathcal{C}_{t,k}\}_{t\in[T],k\in p}}}\sum_{k\in[p]}\left[\max_{\substack{t\in[T]\\ j\in\mathcal{C}_{t,k}}}\Delta_j\right]\left[\max_{\substack{t\in[T]\\ j\in\mathcal{C}_{t,k}}}\frac{20\log(T)}{\Delta_j^2}\right]+5K\right].$$

which implies the bound of the theorem. □

C.2. Regret of UCB-GT

Next, we prove the regret bound for UCB-GT, which demonstrates how one can exploit the bias and feedback structure in the problem.

Proof of Theorem 4.
Proof. As in the previous proof, we focus on bounding $\sum_{t=1}^{T}\mathbb{E}[1_{I_t=i}|\mathbf{G}_t]$ for each arm i. We again split the expectation according to the events based on $Q_{i,t-1}$ as follows:

$$\sum_{t=1}^{T}\mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}\leqslant s_i}|\mathbf{G}_t]+\mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}>s_i}|\mathbf{G}_t],$$

where s_i is to be determined later. We then bound the second term using the algorithm's choice of arm, I_t:

$$\mathbb{E}[1_{I_t=i}1_{Q_{i,t-1}>s_i}|\mathbf{G}_t]=\mathbb{P}[I_t=i,Q_{i,t-1}>s_i|\mathbf{G}_t]\leqslant\mathbb{P}[\widehat{\mu}_{i,t-1}-S_{i,t-1}\leqslant\widehat{\mu}_{*,t-1}-S_{*,t-1},Q_{i,t-1}>s_i|\mathbf{G}_t].$$

$\widehat{\mu}_{i,t-1}$ is a biased estimate of μ_i. This is because whenever x_s falls in the region $\{x\colon r_i(x)>0\wedge r_{I_s}(x)\leqslant 0\}$ and the condition $\widehat{p}_{I_s,i}^{s-1}\leqslant\gamma_{i,s-1}$ holds, the label y_s is not accessible. In this case, the UCB-GT algorithm updates the average loss of expert i optimistically, as if the expert were correct at that time step.

We can decompose this biased estimate $\widehat{\mu}_{i,t-1}$ into two terms: $\widehat{\mu}_{i,t-1}=\widetilde{\mu}_{i,t-1}-\varepsilon_{i,t-1}$. The first term, $\widetilde{\mu}_{i,t-1}$, is an unbiased estimate of arm i and similar to the estimates in Theorem 3. The second term is the misclassification rate $\varepsilon_{i,t-1}$ over $\{s\in[t-1]\colon r_i(x_s)>0\cap r_{I_s}(x_s)\leqslant 0\}$ whenever the condition $\widehat{p}_{I_s,i}^{s-1}\leqslant\gamma_{i,s-1}$ holds, that is, $\varepsilon_{i,t-1}=\frac{1}{Q_{i,t-1}}\sum_{s=1}^{t-1}1_{y_sh_i(x_s)\leqslant 0}1_{r_i(x_s)>0,r_{I_s}(x_s)\leqslant 0}1_{\widehat{p}_{I_s,i}^{s-1}\leqslant\gamma_{i,s-1}}$.

Now, by the design of the UCB-GT, if arm i is chosen at time t, it must be the case that $\widehat{\mu}_{i,t-1}-S_{i,t-1}\leqslant\widehat{\mu}_{*,t-1}-S_{*,t-1}$. We can expand and rewrite this expression as follows:

$$\begin{aligned}&0\leqslant\widehat{\mu}_{*,t-1}+\varepsilon_{i^*,t-1}-\varepsilon_{i^*,t-1}-S_{*,t-1}-\widehat{\mu}_{i,t-1}-\varepsilon_{i,t-1}+\varepsilon_{i,t-1}+S_{i,t-1}\\ \Leftrightarrow\;&0\leqslant(\widetilde{\mu}_{*,t-1}-S_{*,t-1}-\mu_*)+(\mu_i-\widetilde{\mu}_{i,t-1}+S_{i,t-1}-2S_{i,t-1})+(\mu_*-\mu_i+(2+C)S_{i,t-1}),\end{aligned}$$

where we used the fact that $-\varepsilon_{i^*,t-1} \leqslant 0$, and where we bounded $\varepsilon_{i,t-1}$ as follows:

$$\begin{aligned}
\varepsilon_{i,t-1} &= \frac{1}{Q_{i,t-1}} \sum_{s=1}^{t-1} 1_{y_s h_i(x) \leqslant 0} 1_{r_i(x_s)>0, r_{I_s}(x_s) \leqslant 0} 1_{\widehat{p}_{I_s,i}^{s-1} \leqslant \gamma_{i,s-1}} \\
&\leqslant \frac{1}{Q_{i,t-1}} \sum_{s=1}^{t-1} 1_{r_i(x_s)>0, r_{I_s}(x_s) \leqslant 0} 1_{\widehat{p}_{I_s,i}^{s-1} \leqslant \gamma_{i,s-1}} = \frac{1}{Q_{i,t-1}} \sum_{s=1}^{t-1} \sum_{\xi_j \in \mathcal{E}-\xi_i} 1_{r_i(x_s)>0, r_j(x_s) \leqslant 0} 1_{\widehat{p}_{I_s,i}^{s-1} \leqslant \gamma_{i,s-1}} 1_{I_s=j} \\
&\leqslant \frac{1}{Q_{i,t-1}} \sum_{\xi_j \in \mathcal{E}-\xi_i} \sum_{s=1}^{t-1} 1_{r_i(x_s)>0, r_j(x_s) \leqslant 0} 1_{\widehat{p}_{I_s,i}^{s-1} \leqslant \gamma_{i,s-1}}.
\end{aligned}$$

The condition $\widehat{p}_{j,i}^{s-1} \leqslant \gamma_{i,s-1}$ is equivalent to $\sum_{k=1}^{s-1} 1_{r_i(x_k)>0, r_j(x_k) \leqslant 0} \leqslant (s-1)\gamma_{i,s-1}$. Since the sum above is non-zero only when this condition holds, there exists $s_j \in [1, t-1]$ such that $\sum_{s=1}^{t-1} 1_{r_i(x_s)>0, r_j(x_s) \leqslant 0} 1_{\widehat{p}_{j,i}^{s} \leqslant \gamma_{i,s}} \leqslant (s_j-1)\gamma_{i,s_j-1}+1$. Moreover, using the fact that $(s_j-1)\gamma_{i,s_j-1} = \sqrt{5Q_{i,s_j-1}\log(s_j)}/(K-1) \leqslant \sqrt{5Q_{i,t-1}\log(t-1)}/(K-1)$, we can conclude that

$$\varepsilon_{i,t-1} \leqslant \frac{1}{Q_{i,t-1}} \sum_{\xi_j \in \mathcal{E}-\xi_i} \sum_{s=1}^{t-1} 1_{r_i(x_s)>0, r_j(x_s) \leqslant 0} 1_{\widehat{p}_{j,i}^{s-1} \leqslant \gamma_{i,s-1}} \leqslant \frac{K-1}{Q_{i,t-1}} \left[\frac{\sqrt{5Q_{i,t-1}\log(t-1)}}{K-1} + 1 \right] \leqslant C \sqrt{\frac{5\log(t-1)}{Q_{i,t-1}}}$$

for some constant $C > 0$. The rest of the proof now follows by similar arguments as in the proof of Theorem 3. Specifically, we can choose s_i such that the term $\mu_* - \mu_i + (2+C)S_{i,t-1}$ is negative, and since now $\widetilde{\mu}_{*,t-1}$ and $\widetilde{\mu}_{i,t-1}$ are unbiased estimates, we can bound the probabilities $\mathbb{P}[\widetilde{\mu}_{*,t-1} - S_{*,t-1} - \mu_* \geqslant 0 | \mathbf{G}_t]$ and $\mathbb{P}[\mu_i - \widetilde{\mu}_{i,t-1} - S_{i,t-1} \geqslant 0 | \mathbf{G}_t]$ using standard concentration inequalities. □

C.3. Linear regret without the subset property

In this section, we prove Proposition 1, which shows that when the subset property does not hold for a feedback graph, then it is possible to incur linear regret.

Proof of Proposition 1.

Proof. Let $p^* \in (0,1)$. We design a setting in which with probability at least p^*, the UCB-NT algorithm incurs linear regret.

Since the family of abstention functions induces a feedback graph that violates the subset property, there exist pairs (h_i, r_i) and (h_j, r_j) and points x^*, $\widetilde{x}$ for which $x^* \in \mathcal{A}_i \setminus \mathcal{A}_j$, $\widetilde{x} \in \mathcal{A}_i \cap \mathcal{A}_j$, where $\mathcal{A}_i$ and $\mathcal{A}_j$ are the acceptance regions associated with r_i and r_j, respectively, and the feedback graph is designed such that the algorithm updates the pair (h_i, r_i) when the pair (h_j, r_j) is selected.

Now, for some $p \in (0,1)$ to be determined later, consider a distribution with probability p on $(\widetilde{x}, \widetilde{y})$ and $(1-p)$ on (x^*, y^*).

We choose the set of hypothesis functions $\mathcal{H} = \{h_i, h_j\}$, the loss function ℓ in (1), and the labels y^* and $\widetilde{y}$ in such a way that $\ell(\widetilde{y}, h_i(\widetilde{x})) = c - \beta$, $\ell(\widetilde{y}, h_j(\widetilde{x})) = c - \alpha$, and $\ell(y^*, h_i(x^*)) = 0$, where α, β are values that will be later specified. For instance, we can consider the hinge loss $\ell(y, \widehat{y}) = (1 - y\widehat{y})_+$, and h_i, h_j such that $h_i(\widetilde{x}) = \frac{1-c+\beta}{\widetilde{y}}$, $h_j(\widetilde{x}) = \frac{1-c+\alpha}{\widetilde{y}}$, and $h_i(x^*) = \frac{1}{y^*}$. Note that, since $r_j(x^*) < 0$, $\ell(y^*, h_j(x^*))$ may admit any value.

Now, by construction, $\mu_i = (c-\beta)p$ and $\mu_j = (c-\alpha)p + c(1-p) = c - \alpha p$. We claim that we can choose α, β and p such that (1) $\alpha > \beta$; (2) $\mu_i < \mu_j$; (3) $\mu_j < \ell(\widetilde{y}, h_i(\widetilde{x}))$.

The first condition is immediate. The second condition is equivalent to $cp - \beta p < c - \alpha p$, which is itself equivalent to $\alpha - \beta < \frac{c(1-p)}{p}$. By continuity, we can choose α and β close enough such that this is true for any $p \in (0,1)$. The third condition is equivalent to $c - \alpha p < c - \beta$, which is itself equivalent to $\beta < \alpha p$. This is true for p close enough to 1.

Now let $n \in \mathbb{N}$ be large enough such that $\mu_j < \ell(\widetilde{y}, h_i(\widetilde{x})) - \sqrt{\frac{5\log(n)}{n}}$. By continuity, we can choose p large enough such that $p > (p^*)^{1/n}$, and for this choice of p, we can choose α and β such that $\alpha > \beta$, $\alpha, \beta < c$, $\alpha - \beta < \frac{c(1-p)}{p}$, and $\beta < \alpha p$. For instance, if we, without loss of generality, assume that $p > \frac{1}{2}$, then we can choose, $\alpha = \frac{c(1-p)}{2p}$ and $\beta = \frac{c(1-p)}{4}$.

Then, with probability $p^n > p^*$, the point $\widetilde{x}$ will be sampled n times at the start of the game, such that the pair (h_j, r_j) will

Dataset	Number of features
covtype	54
ijcnn	22
skin	3
HIGGS	28
guide	4
phishing	68
cod	8
eye	14
CIFAR	25

Table 1: Table shows the number of features of each dataset.

have a lower confidence bound than the pair (h_i, r_i) at all time steps. Thus, UCB-NT will choose the pair (h_j, r_j) throughout the entire game, even though $\mu_i < \mu_j$. Consequently, the regret of the algorithm will be at least $T(\mu_j - \mu_i)$. □

D. Additional experimental results

In this section, we present several figures showing our experimental results. Figure 7 and Figure 8 show the regret for different abstention costs $c \in \{0.1, 0.2, 0.3\}$ for all our datasets. We observe that, in general, UCB-GT outperforms UCB-NT and UCB for all datasets and is even within the standard deviation of the FS's regret for some of datasets. The figures also indicate that the regret of UCB decreases slowly. This is expected, since there are 2,100 experts, 10,000 time steps, and the algorithm only updates a single expert per time step.

Figure 9 and Figure 10 show the fraction of abstained points for all the datasets. Figure 11 also shows how the fraction of abstained points varies with abstention cost for two extreme values $c \in \{0.001, 0.9\}$. Again UCB-GT admits a lower regret than UCB-NT and UCB and, as expected, the fraction of points decreases as the cost of abstention increases. Figure 12 shows the effect of using confidence-based experts and suggests that the choice of experts does not affect the relative performance of the algorithms. We also tested the effect of varying the number of experts: Figure 13 shows the regret of three datasets when the number of experts is $K = 500$ and $T = 5{,}000$. For this set of experts, we find a similar pattern of performance as above.

Next, we describe in more detail the datasets and how they were processed. In Table 1, we show the number of features for each dataset. For all datasets, we normalized the features to be in the interval $[-1, 1]$. Note that the reason for choosing abstention functions with radius range $(0, \sqrt{d})$ is to cover the entire hypercube $[-1, 1]^d$ with our concentric annuli. For the CIFAR dataset, we extracted the first twenty-five principal components of the horse and boat images, projected the images on these components, and normalized the range of the projections to $[-1, 1]$. The features of the synthetic dataset are drawn from the uniform distribution over $[-1, 1]^2$ and the label is determined by the sign of the projection of a point onto the normal of the diagonal hyperplane $y = -x$.

The confidence-based abstention function has the form $r(x) = |h(x)| - \theta$. In our experiments (Figure 12), we generated twenty abstention functions with thresholds $\theta \in (0, \ldots, 0.25)$, which are paired with each predictor. The predictors are axis-aligned planes along each feature of the dataset. For each dataset, the number of predictors is $\lfloor 100/d \rfloor$ where d is the dimension of the dataset. We chose twenty abstention functions and about 100 prediction functions in order to match the experimental setup of the randomly drawn experts. The total number of experts is then $\lfloor 100/d \rfloor \cdot 20 \cdot d$. Note that we only tested some of our datasets since for larger dimensions d, the number of experts per feature was too small.

D.1. Average regret for different abstention costs and datasets

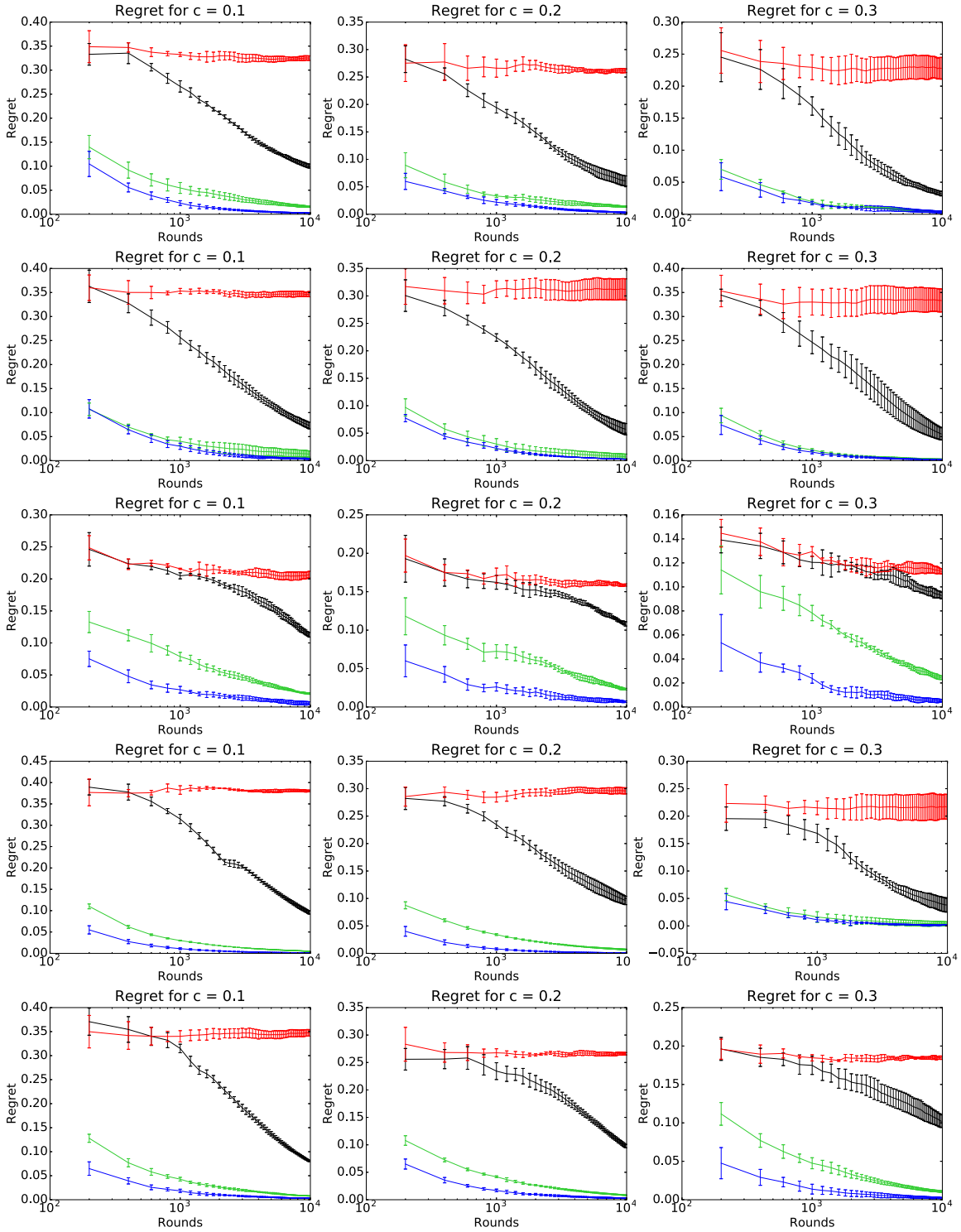

Figure 7: A graph of the averaged regret $R_t(\cdot)/t$ with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. Each row is a dataset, starting from the top row we have: `CIFAR`, `ijcnn`, `HIGGS`, `phishing`, and `covtype`.

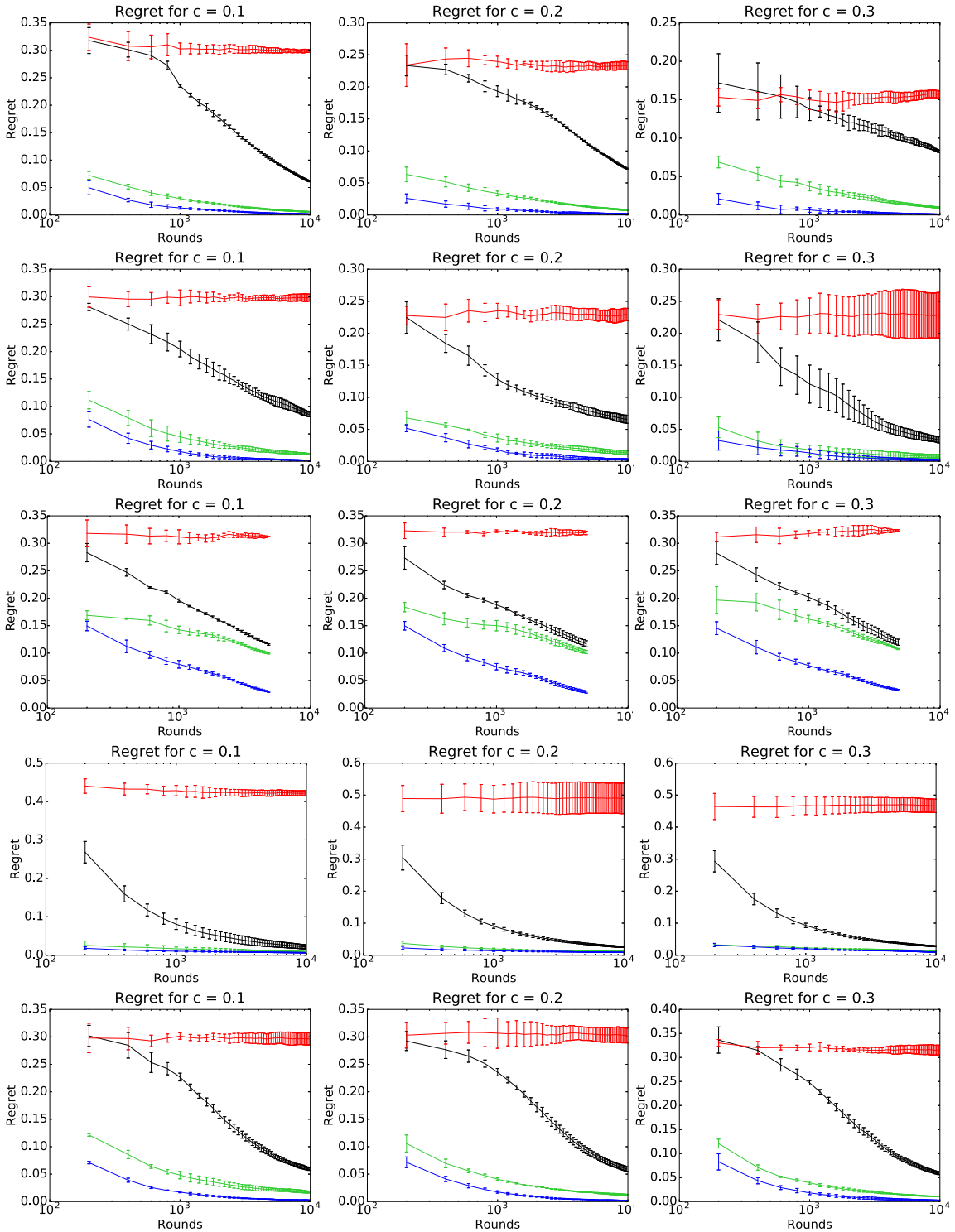

Figure 8: A graph of the averaged regret $R_t(\cdot)/t$ with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. Each row is a dataset, starting from the top row we have: `eye`, `cod-ran`, `synthetic`, `skin`, and `guide`.

D.2. Average fraction of abstention points for different abstention costs and datasets

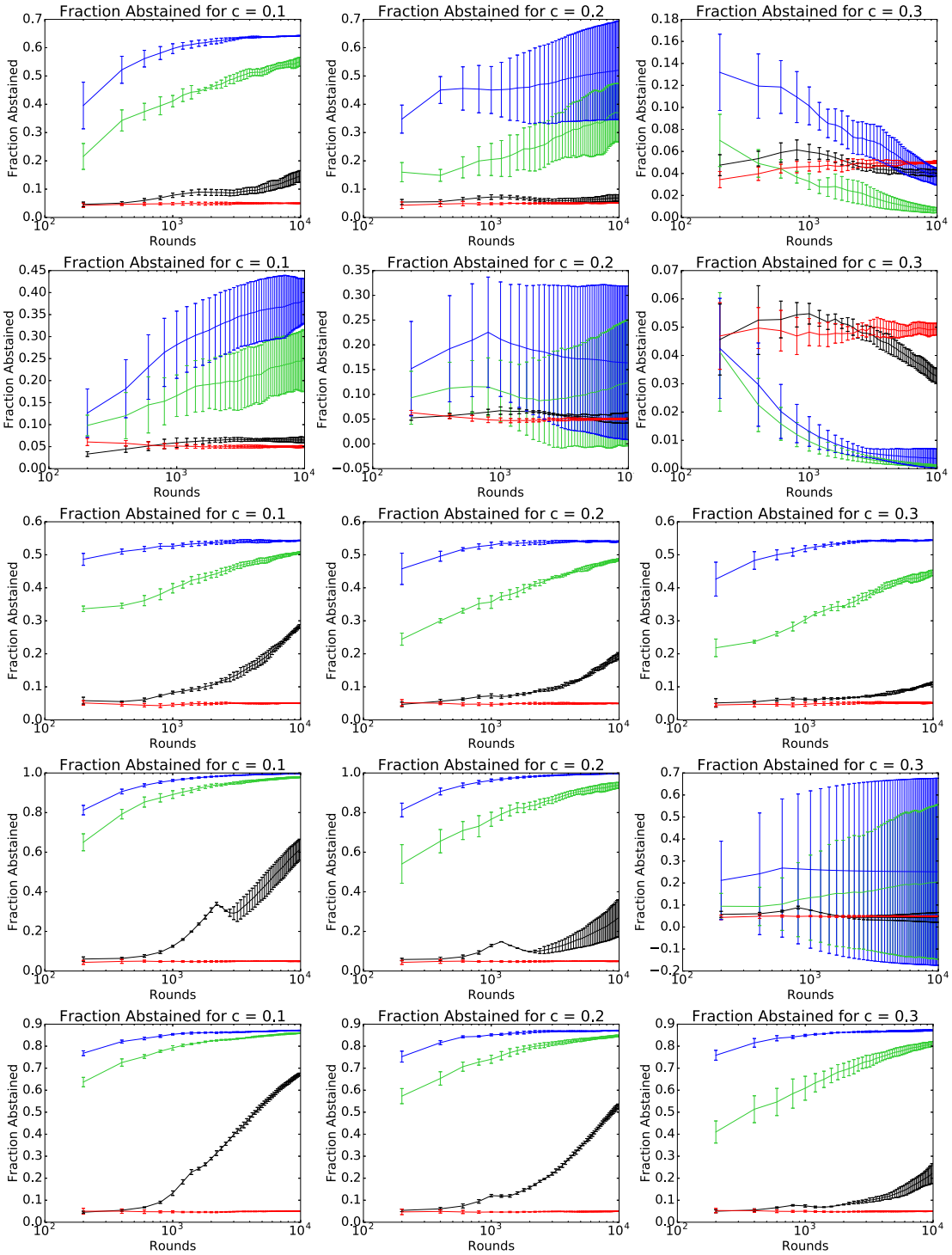

Figure 9: A graph of the averaged fraction of abstained points with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. Each row is a dataset, starting from the top row we have: `CIFAR`, `ijcnn`, `HIGGS`, `phishing`, and `covtype`.

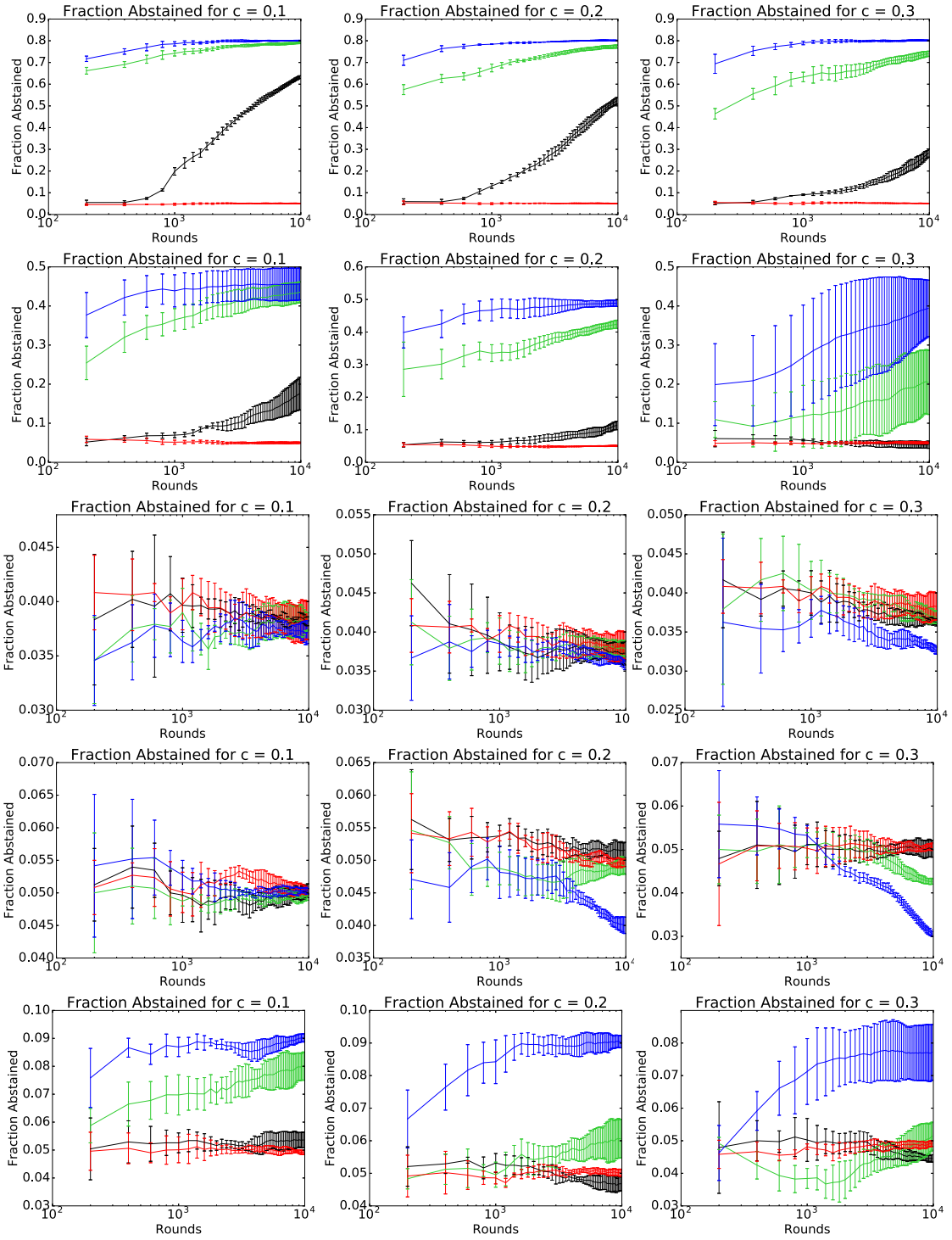

Figure 10: A graph of the averaged fraction of abstained points with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. Each row is a dataset, starting from the top row we have: `eye`, `cod-ran`, `synthetic`, `skin`, and `guide`.

D.3. Average regret and fraction of abstention points for extreme abstention costs

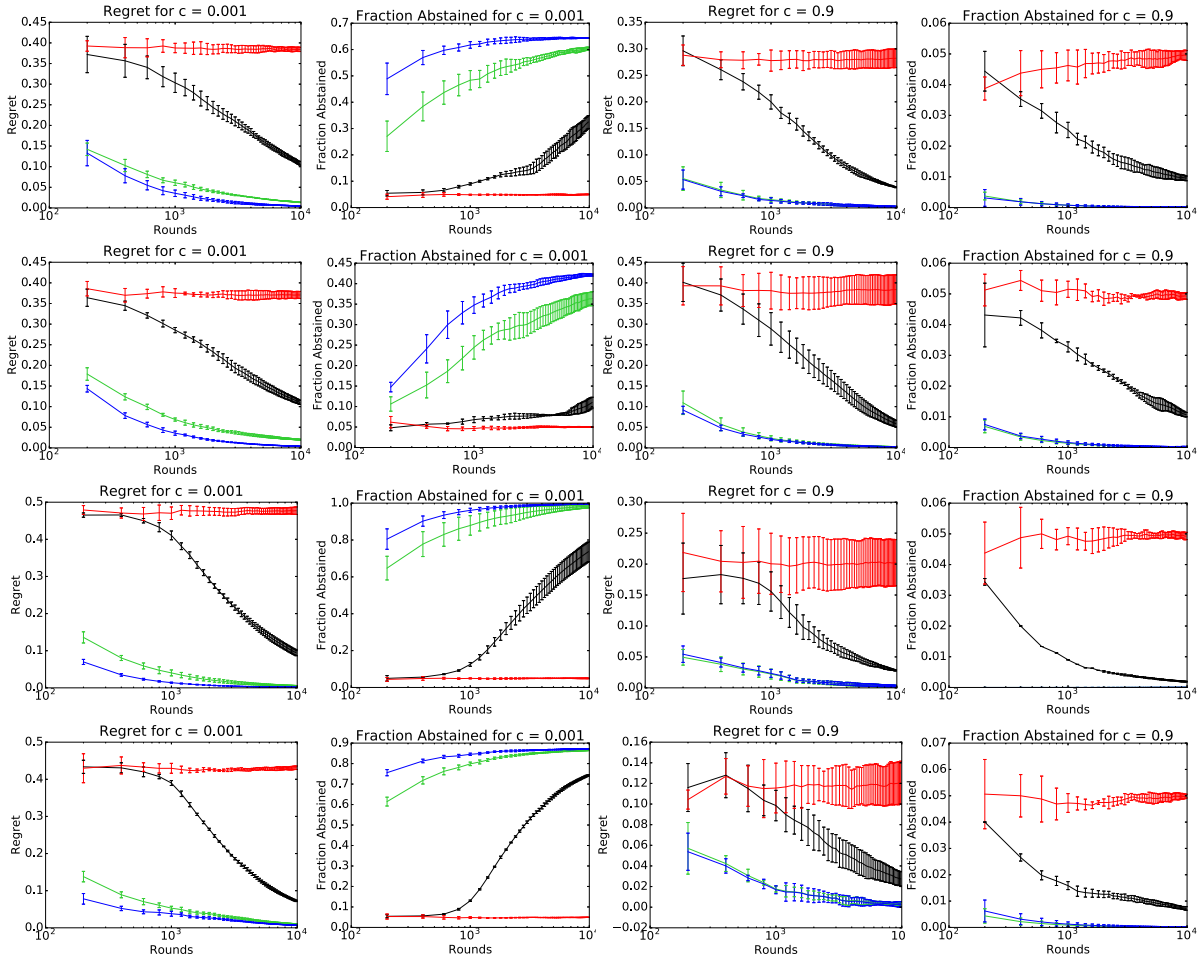

Figure 11: A graph of the averaged regret $R_t(\cdot)/t$ and fraction of points rejected with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. The fraction of points decreases as the cost of abstention increases. The UCB-GT outperforms UCB-NT and UCB while approaching the performance of FS even at these extreme values of c. Each row is a dataset, starting from the top row we have: `CIFAR`, `ijcnn`, `phishing`, and `covtype`.

D.4. Average regret for confidence-based experts

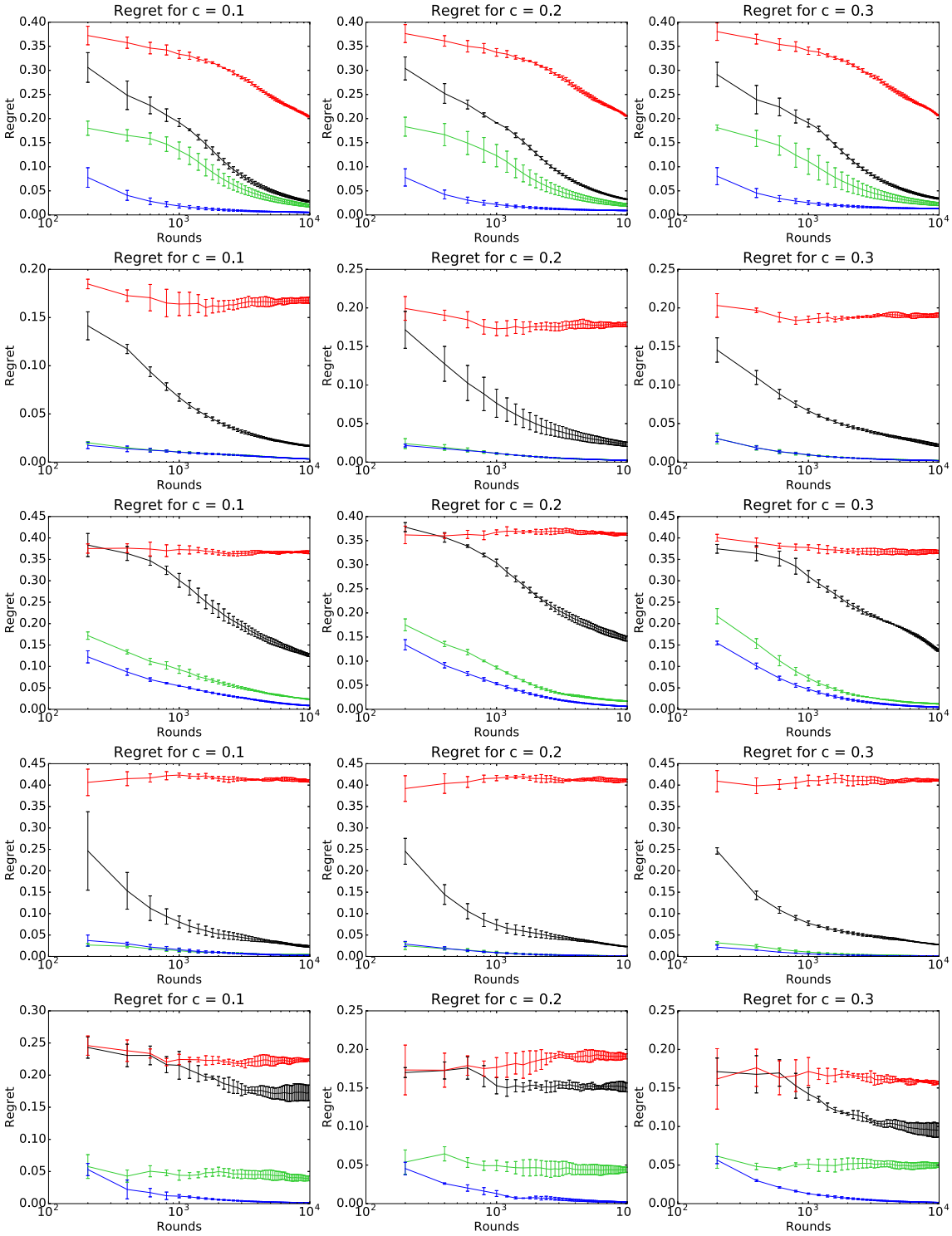

Figure 12: A graph of the averaged regret $R_t(\cdot)/t$ with standard deviations as a function of t (log scale) when using the confidence based experts for UCB-GT, UCB-NT, UCB, and FS. Each row is a dataset, starting from the top row we have: `synthetic`, `skin`, `guide`, `ijcnn` and `CIFAR`.

D.5. Average regret for a smaller set of experts

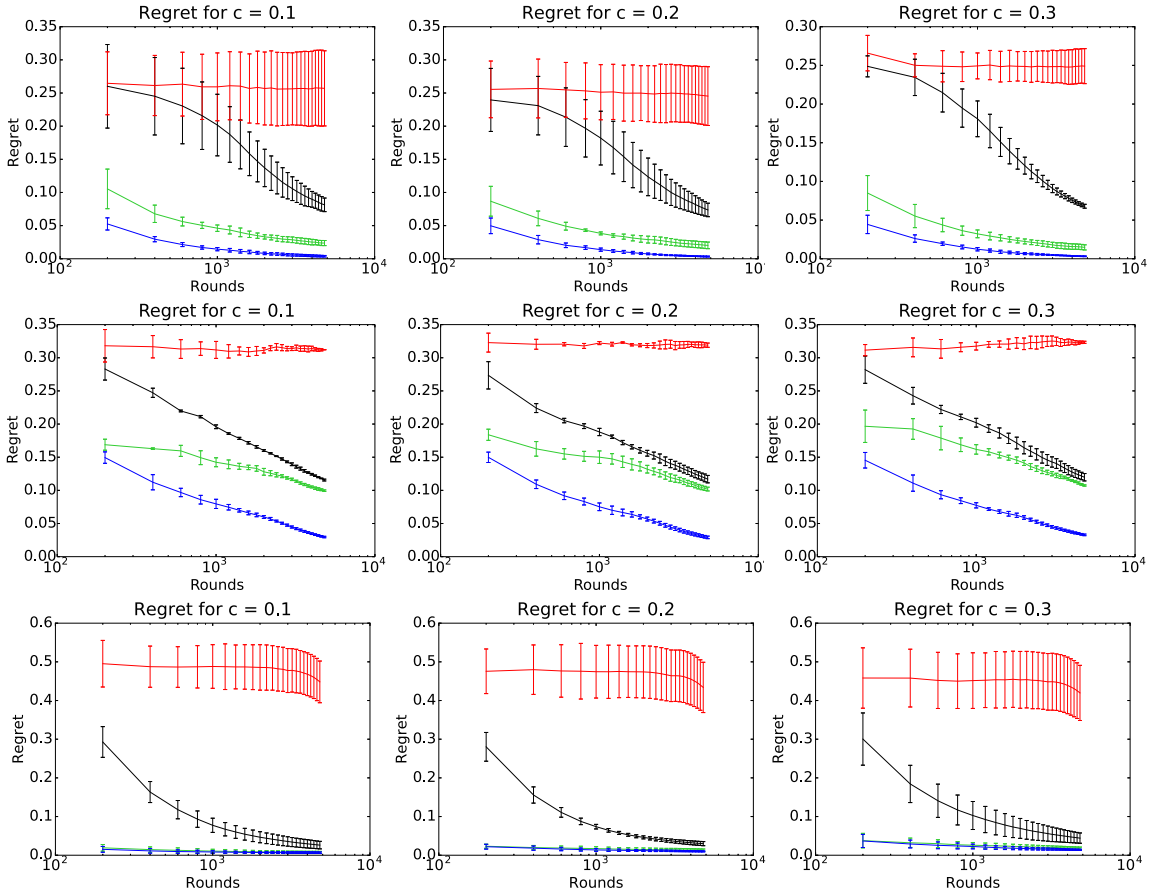

Figure 13: A graph of the averaged regret $R_t(\cdot)/t$ of abstained points with standard deviations as a function of t (log scale) for UCB-GT, UCB-NT, UCB, and FS for different values of abstention costs. Each row is a dataset, starting from the top row we have: `guide`, `synthetic`, and `skin`. We used $K = 500$ experts and $T = 5{,}000$ rounds in order to see the effect when changing the number of experts used.