Online Set Learning from Precision and Recall Feedback

Lee Cohen[*]
Stanford University
leecohencs@gmail.com

Yishay Mansour
Tel Aviv University and Google Research
mansour.yishay@gmail.com

Shay Moran
Technion and Google Research
smoran@technion.ac.il

Han Shao
University of Maryland, College Park
hanshao@umd.edu

Abstract

We consider the problem of learning an unknown subset N^{target} of a domain in an online setting. In each round t, the learner predicts a set of items N_t and receives one of two types of feedback, each with equal probability: *precision feedback*, in which a randomly chosen item from the predicted set N_t is revealed and the learner is told whether it belongs to N^{target} (incurring a reward if it does), or *recall feedback*, in which a randomly chosen item from the target set N^{target} is revealed and the learner is told whether it belongs to N_t (incurring a reward if it does). The goal is to maximize the cumulative reward over time. This simple online set learning problem abstracts a variety of learning scenarios with precision- and recall–type feedback. We show that a hypothesis class (a family of subsets of the domain) is learnable in this setting if and only if it has finite Vapnik-Chervonenkis (VC) dimension, mirroring the classical PAC characterization. However, the resulting algorithmic structure is markedly more intricate: in contrast to standard Probably Approximately Correct (PAC) learning—where the algorithmic landscape is governed by the simple principle of Empirical Risk Minimization (ERM)—our partial feedback model can invalidate ERM and even all proper learning rules. We develop algorithms to address the dependencies induced by the feedback, obtaining regret guarantees in both the realizable and agnostic settings. Our results provide a qualitative characterization of learnability in this model, addressing its most basic question, while pointing to a range of natural and intriguing open questions, including the determination of optimal regret rates.

1 Introduction

In this work, we study the problem of learning an unknown subset N^{target} of a domain $\mathcal{X}$. A basic question in modeling such problems is how to evaluate the quality of a predicted set $\widehat{N} \subseteq \mathcal{X}$. Requiring exact equality between $\widehat{N}$ and N^{target}, as in the classical $0/1$ loss, is often too restrictive: a prediction that captures most elements of N^{target} while including only a few irrelevant items should still be considered accurate.

To capture this notion of approximate correctness, we adopt the standard measures of *precision* and *recall*. Precision is the fraction of predicted elements that belong to N^{target}, while recall is the fraction of N^{target} that appears in the prediction. These measures are standard in problems where predictions are sets and quality depends on both relevance and coverage, including information

[*]Authors are ordered alphabetically.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

retrieval (Manning et al., 2008), data mining (Han et al., 2011), and other subset-prediction tasks. We remark that in the statistics literature, recall is also called *sensitivity* or *True Positive Rate (TPR)*, precision is also called *Positive Predictive Value (PPV)*.

We study this problem in a *sequential online learning setting*. At each round $t = 1, \ldots, T$, the learner outputs a subset $N_t \subseteq \mathcal{X}$. Then, one of two feedback types is selected uniformly at random: with probability $1/2$ the learner receives *precision feedback*, and with probability $1/2$ it receives *recall feedback*. In the precision case, a uniformly random element is drawn from the predicted set N_t, and the learner is given this element along with an indication of whether it belongs to N^{target} (receiving a reward of 1 if it does). In the recall case, a uniformly random element is drawn from the target set N^{target}, and the learner is given this element along with an indication of whether it belongs to N_t (receiving a reward of 1 if it does). Thus, each round provides information about a single element, sampled either from the prediction or from the target set. Precision feedback reflects the relevance of predicted items, while recall feedback reflects coverage of the target set. The learner must use this limited, sequential feedback to refine its predictions over time.

The goal is to minimize *regret*: we assume a hypothesis class $\mathcal{H} \subseteq 2^{\mathcal{X}}$ of candidate subsets is given, and define regret as the gap between the learner's cumulative reward and that of the best fixed $N \in \mathcal{H}$ in hindsight. We consider two standard settings. In the *realizable* setting, there exists $N \in \mathcal{H}$ that achieves perfect precision and recall with respect to N^{target}. In the *agnostic* setting, no such assumption is made, and the learner competes with the best $N \in \mathcal{H}$.

The model can be viewed as an abstraction of repeated decision problems in which a system proposes a set of items and receives only item-level feedback. For example, in content recommendation, a platform presents a list of items, and feedback may arise either from an item the user selects from the list (precision-type feedback) or from an item the user seeks out independently (recall-type feedback). A related scenario arises in hiring. The goal of an employer or HR team is to identify strong candidates for a role, and they may proactively reach out to a set of candidates they believe are promising. Precision feedback arises when a candidate from this list is evaluated—for instance, a successful hire from the list provides positive feedback, while an unsuccessful hire provides negative feedback. At the same time, candidates may apply independently, outside the proposed list; when such a candidate turns out to be a successful hire, this provides recall-type feedback, indicating that a relevant element was missed. Similar patterns appear in other domains where decisions are made over sets of candidates. For example, in cybersecurity, a detection system might output a set of IP addresses that seem suspicious. A security analyst may inspect a flagged IP address and determine whether it is malicious, yielding precision feedback. Alternatively, an attack may be observed from a malicious IP address that was not necessarily flagged, yielding recall feedback depending on whether that IP was included in the predicted set.

These examples are intended mainly to build intuition. Our goal is not to faithfully model any specific application, but rather to isolate a basic theoretical learning problem: learning sets from item-level precision and recall feedback.

Our Contributions. Our main result is a complete characterization of learnability: *a hypothesis class $\mathcal{H}$ is learnable in this setting if and only if it has finite VC dimension.* This parallels the classical PAC characterization, where VC dimension governs learnability. At the same time, the learning problem here is structurally different: the feedback depends on the learner's prediction and is obtained by sampling either from the predicted set or from the target set. This induces two complementary but asymmetric signals - precision (relevance) and recall (coverage) - which leads to algorithmic behavior that differs substantially from both supervised learning and classical bandit models.

Concretely, we show:

- In the realizable setting, we propose an algorithm that selects the smallest hypothesis consistent with the observed recall feedback, and prove that it achieves regret $O(d \log^2 T)$, where d is the VC dimension of $\mathcal{H}$. We also show that any algorithm must incur regret at least $\Omega(d)$.
- In the agnostic setting, we provide an algorithm achieving regret $\widetilde{O}(d^{1/4}T^{3/4})$, showing that sublinear regret is achievable even without realizability assumptions.
- We identify several phenomena that distinguish this setting from classical learning models. In the realizable case, selecting an arbitrary hypothesis in $\mathcal{H}$ that is consistent with all feedback observed so far (as in empirical risk minimization) does not, in general, yield a sound learner. However, a

regularized variant of ERM that prioritizes the *smallest* consistent hypothesis in the class yields vanishing average loss. In contrast, in the agnostic setting, proper learning may fail: achieving low regret sometimes requires using hypotheses outside the class $\mathcal{H}$.

We also establish a connection to online learning: any online classification algorithm with a finite mistake bound can be transformed into an algorithm for our setting with a comparable mistake bound. This connection is one-way: there exist hypothesis classes that are not online learnable, yet are learnable in our setting. This separation arises because learnability in our model is characterized by VC dimension, whereas online learnability is governed by the Littlestone dimension, which can be arbitrarily larger.

Finally, in the realizable case, we show that recall feedback alone suffices for learning, and that precision feedback is not necessary. This phenomenon appears to be specific to the realizable setting; in particular, our algorithm for the agnostic case relies on both types of feedback.

Overall, even this basic set-learning model exhibits a range of non-trivial learning behaviors. Our results provide a complete characterization of learnability in this setting via VC dimension. A natural next step is to refine this understanding by determining optimal regret and mistake bounds for specific hypothesis classes. In particular, it would be interesting to understand whether the agnostic regret can be improved to the more familiar $\sqrt{dT}$ rate, and whether the $\log^2 T$ factor in the realizable case can be removed. We believe these questions present a number of interesting and challenging problems.

Related Work. Precision and recall are natural and standard metrics used broadly in machine learning, spanning applications from binary classification Juba and Le (2019), multi-class classification Grandini et al. (2020), regression Torgo and Ribeiro (2009), and time series Tatbul et al. (2018) to information retrieval Arora et al. (2016) and generative models Sajjadi et al. (2018). Beyond precision-recall, another related metric—the area under the ROC curve (AUC)—has also been extensively studied in the history of binary classification Cortes and Mohri (2003, 2004); Rosset (2004); Agarwal et al. (2005), with a focus on generalization.

Multi-label learning McCallum (1999); Schapire and Singer (2000) has been an area of study in machine learning, with various, primarily experimental approaches (see, e.g., Elisseeff and Weston (2001); Petterson and Caetano (2011); Kapoor et al. (2012) and Zhang and Zhou (2014); Bogatinovski et al. (2022) for surveys). In multi-label learning, the training set consists of examples, each associated with multiple labels rather than just one. The goal is to train a model that can learn the relationships between the features of each example and all its labels. At test time, the learner predicts a list of labels for new examples, aiming to capture all the relevant labels, rather than just a single one.

Some works have studied Bayes-consistency of surrogate losses in multi-label learning, where direct loss minimization is often infeasible Gao and Zhou (2011). Extensions include rank-based metrics like precision@κ and recall@κ (Menon et al., 2019), which are loss functions defined under the constraint that the number of labels predicted by the model is limited to κ. Recently, Mao et al. (2024) provided stronger H-consistency bounds with non-asymptotic guarantees. Unlike the fully supervised multi-label formulation with complete per-example label sets, we consider an online learning problem and our goal is to identify a single hidden subset using only random positive samples and bandit-style precision feedback[2], without observing the entire target set. Precision feedback can reveal negative labels, whereas recall feedback supplies only positive examples. A recent work Cohen et al. (2024) considered precision and recall learning in the PAC model for user-dependent sets, and derived guarantees under recall-only feedback; in contrast, our results deliver online regret bounds for precision and recall in a partial-feedback setting.

Positive-unlabeled learning provides a related perspective on recall feedback. In particular, Liu et al. (2002) analyze a rule that covers observed positive examples while minimizing the mass classified as positive, a principle related to our smallest-consistent-set rule. Mansouri and Ben-David (2025) establish finite-sample guarantees under several positive-sampling assumptions, including settings without a known class prior. These works study classification from positive and unlabeled samples; our learner knows the finite available domain and is evaluated by cumulative precision/recall reward, with additional prediction-dependent precision feedback. Another related line studies learning with multiple correct answers. Joshi et al. (2026) learn to produce a good answer from demonstrations, with complexity controlled by a reward class, and Pour et al. (2026) compare online guarantees when

[2]Bandit-style means precision is only observable for the implemented hypothesis: given $N_t = \{A, B, C\}$, feedback comes from a sampled item in N_t and cannot evaluate other sets such as $\{D, E\}$. Thus, feedback is one-item and hypothesis-dependent.

feedback reveals a correct answer, a correct answer together with prediction correctness, or the full acceptable set. Those formulations seek a single acceptable answer for each context, whereas our predictions are sets and the objective rewards both relevance and coverage.

2 Model

In our model, we distinguish between a possibly infinite universe of items $\mathcal{X}$, on which the hypothesis class $\mathcal{H} \subseteq 2^{\mathcal{X}}$ is defined, and a fixed, finite available set $X \subseteq \mathcal{X}$ in a given learning instance. The learner *knows* X. Both $\mathcal{X}$ and $\mathcal{H}$ may be infinite, but predictions and feedback are restricted to the finite set X. For intuition, one may think of $\mathcal{X}$ as the set of all dishes a restaurant could potentially offer, while X is the current menu, determined by seasonality, ingredients, or the chef's daily choices. The learning task could then be to learn a personalized submenu for a customer based on the currently available dishes in X. This distinction allows us to define learnability in a qualitative, "distribution-free" manner — uniformly over all finite sets $X \subseteq \mathcal{X}$ — even for infinite hypothesis classes over infinite domains $\mathcal{X}$. In particular, it clarifies the role of VC dimension as the quantity characterizing learnability. See Appendix C for a discussion on $X = \mathcal{X}$ versus $X \subseteq \mathcal{X}$.

Let $N^{\text{target}} \subseteq X$ be the *unknown* target set. We consider an online setting over T rounds. At each round $t = 1, \ldots, T$, the learner selects a predicted set $N_t \subseteq X$, and one of the following two feedback types is selected with equal probability (i.e., probability $1/2$):

- **Recall feedback:** a uniformly random item $u_t \sim \text{Unif}(N^{\text{target}})$ is sampled, and the learner observes it and receives recall reward $g_t(N_t) = \mathbb{1}(u_t \in N_t)$.
- **Precision feedback:** a uniformly random item $v_t \sim \text{Unif}(N_t)$ is sampled, along with a label $y_t = \mathbb{1}(v_t \in N^{\text{target}})$, and the learner observes the item and receives precision reward $g_t(N_t) = y_t$.

The *(expected) reward* of a set $N \subseteq X$ is defined as:

$$g(N) = \frac{1}{2} \underbrace{\mathbb{P}_{u \sim \text{Unif}(N^{\text{target}})} \left(u \in N\right)}_{\text{recall:} r(N)} + \frac{1}{2} \underbrace{\mathbb{P}_{v \sim \text{Unif}(N)} \left(v \in N^{\text{target}}\right)}_{\text{precision:} p(N)} .$$

The first term, recall $r(N)$, is the probability that a uniformly random target item $u \sim \text{Unif}(N^{\text{target}})$ is included in N. The second term, precision $p(N)$, is the probability that a uniformly random item $v \sim \text{Unif}(N)$ belongs to N^{target}.

Recall can also be expressed as $r(N) = \frac{|N \cap N^{\text{target}}|}{|N^{\text{target}}|}$, which is the fraction of items in N^{target} that also belong to N. Similarly, precision can be expressed as $p(N) = \frac{|N \cap N^{\text{target}}|}{|N|}$, which is the fraction of items in N that also belong to N^{target}.[3]

Given any universal set $N \subseteq \mathcal{X}$ (which may include items outside X), the set effectively available to the learner is its restriction, $N_{|X} = N \cap X$. For example, if $N = N^{\text{target}} \cup N^{\text{unavailable}}$, where N^{target} is the target set and $N^{\text{unavailable}} \subseteq \mathcal{X} \setminus X$ consists of some unavailable items, then none of the items in $N^{\text{unavailable}}$ can be selected. Consequently, the set of items N coincides exactly with N^{target}. In this case, N achieves perfect precision and recall, and therefore receives the same reward as N^{target}. We therefore generalize the definition of reward to any $N \subseteq \mathcal{X}$: if N contains items outside X, only the subset $N_{|X}$ can be played. Accordingly, we define the reward of any set N as

$$g(N) = g(N_{|X}).$$

Then given a hypothesis class $\mathcal{H} \subseteq 2^{\mathcal{X}}$, the *optimal* hypothesis is

$$N^\star = \arg\max_{N \in \mathcal{H}} g(N) .$$

The learner's goal is to minimize the regret, defined in terms of expected rewards as

$$\text{Reg}_T = \sum_{t=1}^{T} \big(g(N^\star) - g(N_t)\big) .$$

[3] We set $r(N) = 1$ when $N^{\text{target}} = \emptyset$ and $p(N) = 1$ when $N = \emptyset$. Sampling from an empty source returns a null symbol and reward 1 for that feedback type, rather than an item. Thus null recall feedback reveals an empty target.

This quantity is random because the predictions depend on the observed feedback and any internal randomization. Its expectation satisfies $\mathbb{E}[\mathrm{Reg}_T] = Tg(N^\star) - \mathbb{E}\left[\sum_{t=1}^{T} g_t(N_t)\right]$. We say a hypothesis class $\mathcal{H}$ is *learnable* if we can always achieve diminishing regret for any X and N^{target}. More formally,

Definition 2.1 (Learnability)**.** A hypothesis class $\mathcal{H}$ is learnable if there exists a learning rule and a function $R(T, \delta)$ such that, for every fixed $\delta \in (0, 1)$, $R(T, \delta) = o(T)$, and for every finite available set $X \subseteq \mathcal{X}$, target set $N^{\text{target}} \subseteq X$, and horizon T, the learning rule (given X) achieves $\mathrm{Reg}_T \leq R(T, \delta)$ with probability at least $1 - \delta$. The same bound must hold uniformly over all such X and N^{target}; it may depend on $\mathcal{H}$, but not on the size of the available set.

This raises the central question:

Which hypothesis classes are learnable?

Note that our learning problem is not full-information. While we can observe a recall reward for each hypothesis $N \in \mathcal{H}$, i.e., $\mathbb{1}(u_t \in N)$ after receiving the recall feedback $u_t \sim \mathrm{Unif}(N^{\text{target}})$, the situation is different for precision. In precision feedback, the item v_t is sampled from the predicted list, so we can only observe a precision reward for the selected hypothesis, not for all hypotheses in $\mathcal{H}$. *Thus, recall feedback is full-information-style (i.e., evaluable across all hypotheses), whereas precision feedback is bandit-style (i.e., limited to the selected hypothesis).* For a finite hypothesis class, a standard bandit algorithm such as EXP3 yields the expected-regret bound:

$$\mathbb{E}[\mathrm{Reg}_T] = O\left(\sqrt{T|\mathcal{H}|\log|\mathcal{H}|}\right),$$

which scales polynomially with the size of the hypothesis class $|\mathcal{H}|$. Since this proper learner has nonnegative pseudo-regret, Markov's inequality also gives a sublinear bound for every fixed failure probability. Thus, any finite class is learnable. However, when the class is large, i.e., $|\mathcal{H}| \geq T$, this regret bound becomes vacuous, and it remains unclear which infinite classes are learnable. On a fixed finite pool, one can also run EXP3 over the distinct restrictions $\mathcal{H}_{|X} = \{N \cap X : N \in \mathcal{H}\}$ together with $\emptyset$. When $\emptyset \notin \mathcal{H}_{|X}$, this adds an improper action, so a lower bound restricted to proper learners does not rule out this approach. The resulting guarantee still depends on the number of actions, rather than only on VC dimension. A standard cover under a fixed reference distribution does not immediately resolve this dependence, because precision is evaluated under the prediction-dependent distribution $\mathrm{Unif}(N)$.

In the remainder of the paper, we simplify the setting by assuming that in each round we observe both precision and recall feedback. This assumption is justified as follows: in the original setting, where only one type of feedback (precision or recall) is observed per round at random, we can simulate both types by repeatedly using the same hypothesis for $O(\log(T/\delta))$ rounds. With high probability (at least $1 - \delta$), this ensures receiving both feedback types for each implemented hypothesis. As a result, any algorithm designed for the simplified setting with regret bound R can be extended to the original setting with regret at most $O(\log(T/\delta)) \cdot R$. Any regret lower bound in this simplified setting is also a lower bound in the original setting.

Each set $N \subseteq \mathcal{X}$ naturally defines a binary classifier $h_N : \mathcal{X} \to \{0, 1\}$ via $h_N(x) = \mathbb{1}(x \in N)$. We define the VC dimension and Littlestone dimension of the hypothesis class $\mathcal{H}$ as the VC and Littlestone dimensions of the corresponding class of binary functions $\{h_N \mid N \in \mathcal{H}\}$. We now present regret bounds in terms of these dimensions.

3 Realizable Setting

In the realizable setting, we assume that the optimal set $N^\star$ incurs no loss, i.e., $g(N^\star) = 1$. We propose two algorithmic approaches: one based on the mistake-bound model from online learning, and the other inspired by a reduction from PAC learning to stochastic online learning, where in each round an Empirical Risk Minimizer is selected based on the historical data. The first approach yields a regret guarantee based on the Littlestone dimension, while the second yields a regret bound of $O(\mathrm{VCdim}(\mathcal{H}) \log^2 T)$ based on the VC dimension. The former is quantitatively better when the hypothesis class has finite Littlestone dimension. We delegate all missing proofs to the appendix.

Algorithm 1 Reduction to Online Classification

Input: standard online learning algorithm $\mathcal{A}$
for $t = 1, \ldots, T$ **do**
  Let $N_t = \{x | \mathcal{A} \text{ predicts positive at } x\}$
  if observe precision feedback v_t and make a mistake $v_t \notin N^{\text{target}}$ **then**
    Feed $\mathcal{A}$ with $(v_t, 0)$.
  else
    Feed $\mathcal{A}$ with $(u_t, 1)$, where u_t is a recall feedback.
  end if
end for

3.1 Reduction to Standard Online Learning

Any standard online learning algorithm $\mathcal{A}$ for binary classification can be applied directly here (Algorithm 1) and incurs the same amount of mistakes.

Theorem 3.1. *For any $X \subseteq \mathcal{X}$ and $N^{target} \subseteq X$, for any realizable hypothesis class $\mathcal{H}$, for any online classification algorithm $\mathcal{A}$ with mistake bound M for $\mathcal{H}$, Algorithm 1 will incur regret $Reg_T \leq M$. By choosing $\mathcal{A}$ to be the standard optimal algorithm by Littlestone (1988), Algorithm 1 will incur regret $Reg_T \leq \mathrm{Ldim}(\mathcal{H})$, where $\mathrm{Ldim}(\mathcal{H})$ is the Littlestone dimension of $\mathcal{H}$.*

Hence, any hypothesis class with finite Littlestone dimension is learnable in our setting. However, in the following, we show that the Littlestone dimension is not the appropriate complexity measure in our setting; instead, the VC dimension is the relevant one.

3.2 Hypothesis Classes with Finite VC Dimension are Learnable

We now present a different algorithm inspired by Empirical Risk Minimization (ERM). In standard online/i.i.d. realizable setting, selecting any empirical risk minimizer—i.e., an arbitrary hypothesis that is consistent with the observed history—yields a regret bound of $O(\mathrm{VCdim}(\mathcal{H}) \log T)$. In our setting, we can also achieve a regret of $O(\mathrm{VCdim}(\mathcal{H}) \log^2 T)$ using a similar idea: at round t, let $H_t = \{N \in \mathcal{H} \mid u_\tau \in N, \forall \tau \in [t-1]\}$ denote the set of hypotheses consistent with all historical *recall* feedback. Among them, we choose a hypothesis with a minimal size (intersecting with the available items X),

$$N_t \in \arg\min_{N \in H_t} |N \cap X| ,$$

where $|N \cap X|$ is the number of available items predicted by N. This corresponds to selecting the hypothesis with maximum likelihood.

Theorem 3.2. *For any $X \subseteq \mathcal{X}$ and $N^{target} \subseteq X$, for any realizable hypothesis class $\mathcal{H}$, choosing the maximum likelihood hypothesis in each round achieves regret of $(d \log(T) + \log(T/\delta)) \log T$ with probability at least $1 - \delta$, where $d = \mathrm{VCdim}(\mathcal{H})$.*

Note that the maximum likelihood predictor is a specific hypothesis consistent with all historical recall feedback and does not utilize the precision feedback at all. A natural question raised here is: *would any arbitrary hypothesis consistent with recall feedback work?* The answer is no. Consider an example of input space $\mathcal{X} = \{0, 1, \ldots, n\}$ and hypothesis class $\mathcal{H} = \{\{0\}, \{0, 1\}, \ldots, \{0, n\}\}$. When $X = \mathcal{X}$ and $N^{\text{target}} = \{0\}$, all hypotheses in $\mathcal{H}$ will be consistent with the recall feedback as we will only observe $u_t = 0$. But all the other hypotheses, e.g., $\{0, 1\}$ will incur $\frac{1}{2}$ loss in precision. This means that, in general, ERMs do not work for this problem. Next, we show a lower bound that depends on the $\mathrm{VCdim}(\mathcal{H})$.

Theorem 3.3 (Lower bound)**.** *For any $d > 0$ and any hypothesis class $\mathcal{H}$ with $\mathrm{VCdim}(\mathcal{H}) = d$, there exists an available set X such that for any (possibly randomized) online learner, there exists a target set N^{target} realizable by $\mathcal{H}$ for which the learner will suffer expected regret $\mathbb{E}[Reg_T] \geq d/50$ for $T \geq d/10$.*

There is a gap of $\log^2(T)$ between the upper and lower bounds. However, we show that it is impossible to improve the lower bound to $d \log^2(T)$. The optimal regret bound is left as an open question.

Proposition 3.4. *For any $d > 0$, there exists a hypothesis class $\mathcal{H}$ with $VCdim(\mathcal{H}) = \log |\mathcal{H}| = d$ such that for any $X \subseteq \mathcal{X}$, we can achieve regret $Reg_T \leq d$.*

4 Agnostic Setting

In the agnostic setting, the reward of the optimal hypothesis, $g(N^\star) = \max_{N \in \mathcal{H}} g(N)$, does not necessarily equal 1.

4.1 Impossibility Result for Proper Algorithms

A learning algorithm is said to be *proper* if, in each round, it chooses a hypothesis from the hypothesis class $\mathcal{H}$. Most well-known online learning algorithms—such as Multiplicative Weights, Follow the Perturbed Leader, and EXP3—as well as our algorithms for the realizable setting, are proper. However, in the agnostic setting, proper learning rules do not work in general.

Specifically, we establish the following impossibility result.

Theorem 4.1. *There exists a hypothesis class $\mathcal{H}$ with $\mathrm{VCdim}(\mathcal{H}) = 1$ such that, for any T, there exists an available set X and target set $N^{target} \subseteq X$ for which any proper learning algorithm incurs expected regret $\mathbb{E}[Reg_T] \geq C \cdot T$ for some universal numerical constant $C > 0$.*

At a high level, in this construction, all hypotheses in $\mathcal{H}$ are large and mutually disjoint. The target set selects T hypotheses, and then includes a $\frac{1}{2} + \varepsilon$ fraction of one of them, and a $\frac{1}{2}$ fraction from each of the remaining $T - 1$ hypotheses. This effectively reduces the problem to a bandit instance with T independent arms.

However, readers might observe that the reward of every hypothesis in $\mathcal{H}$ is quite low–roughly $\frac{1}{4}$–since the recall is approximately $\frac{1}{T}$ and the precision is around $\frac{1}{2}$. Notably, both the empty set $\emptyset$ and the full available set X achieve a higher reward, namely $g(X), g(\emptyset) \geq \frac{1}{2}$. It follows that whenever the best hypothesis in the class satisfies $g(N^\star) < \frac{1}{2}$, one can outperform it by simply outputting either $\emptyset$ or X. This observation motivates a natural improper learning rule: output the empty set $\emptyset$ or the full available set X whenever it is detected that no hypothesis in $\mathcal{H}$ performs well. Based on this improper algorithmic principle, we propose a new algorithm in the following.

4.2 The Algorithm

We define the optimality margin of $\mathcal{H}$ as follows.

Definition 4.2 (Optimality margin)**.** Given a hypothesis class $\mathcal{H}$ and available set X, let $N^\star \in \arg\max_{N \in \mathcal{H}} g(N)$ be any maximizer of the expected reward. The *optimality margin* of $\mathcal{H}$ is defined as

$$\alpha_{\mathcal{H}} = 2g(N^\star) - 1,$$

which measures twice the difference between the optimal reward and the baseline $1/2$. For a nonempty target, $g(\emptyset) = 1/2$, whereas $g(X) = (1 + p(X))/2 \geq 1/2$ and can be strictly larger. If the target is empty, the empty prediction has reward 1 under our conventions.

Since $g(N^\star) = \frac{1}{2}(r(N^\star) + p(N^\star))$ and $g(N^\star) = \frac{1}{2}\alpha_{\mathcal{H}} + \frac{1}{2}$, we have $r(N^\star) + p(N^\star) = 1 + \alpha_{\mathcal{H}}$. Since both precision and recall are at most 1, the optimal hypothesis $N^\star$ has both $r(N^\star) \geq \alpha_{\mathcal{H}}$ and $p(N^\star) \geq \alpha_{\mathcal{H}}$.

We construct APRIL (Algorithm 2), an improper algorithm that compares the surviving hypotheses with the empty prediction, whose reward is $1/2$ for a nonempty target. It achieves regret $\widetilde{O}(d^{1/4}T^{3/4})$ without knowing $\alpha_{\mathcal{H}}$. If a positive lower bound on $\alpha_{\mathcal{H}}$ is supplied, its guarantee improves to a rate of order $\sqrt{T}$, up to the dimension, confidence, and exploration terms below. Throughout the algorithm and its analysis, we identify each hypothesis, including $N^\star$, with its restriction to X.

Precision feedback is observed only for the predicted set. To estimate precision across a candidate class, we use $|N^{\text{target}}| = p(N)|N \cap X|/r(N)$ whenever $r(N) > 0$. Given an empirical recall function $\widehat{r}$ and m nonempty predictions with precision labels, define

$$\widehat{n}^\star = \mathrm{EST}(\widehat{r}, (N_t, y_t)_{t=1}^m) = \frac{1}{m} \sum_{t=1}^{m} \frac{y_t |N_t \cap X|}{\widehat{r}(N_t)}. \tag{1}$$

The algorithm uses this estimate only when $m \geq 1$ and all its denominators are positive. Recall and precision lower bounds on the sampled hypotheses allow us to control the estimator despite adaptive predictions.

Algorithm 2 APRIL: Adaptive Precision-Recall Interactive Learner

Input: $X, \mathcal{H}, T, d = \mathrm{VCdim}(\mathcal{H})$, $\alpha \in (0, 1]$, $\delta \in (0, 1)$.
Stage 1: Predict X for

$$T_1 = \left\lceil \frac{1024}{\alpha^2} \left(d \log \frac{64d}{\alpha^2} + \log \frac{32}{\delta} \right) \right\rceil$$

rounds and set $\widehat{r}_0(N) = T_1^{-1} \sum_{t=1}^{T_1} \mathbb{1}(u_t \in N)$.
Set $H_1 = \{N \in \mathcal{H}_{|X} : \widehat{r}_0(N) \geq 7\alpha/8\}$.
Stage 2: Choose $N_1 \in \arg\max_{N \in H_1} \widehat{r}_0(N)/|N \cap X|$ and predict it for $T_2 = \lceil 64 \log(8/\delta)/\alpha^2 \rceil$ rounds.
Set $\widehat{n}_0^\star = \mathrm{EST}(\widehat{r}_0, (N_1, y_t)_{t=1}^{T_2})$ and $\widehat{p}_0(N) = \widehat{r}_0(N)\widehat{n}_0^\star/|N \cap X|$.
Set $\widehat{\mathcal{H}} = \{N \in H_1 : \widehat{p}_0(N) \geq 7\alpha/16\}$.
Stage 3: Restart the observation history and its indices.
for $t = 1, \ldots, T_3 = T - T_1 - T_2$ **do**
  Update $\widehat{r}_t(N) = \frac{1}{t-1} \sum_{\tau=1}^{t-1} \mathbb{1}(u_\tau \in N)$, $\widehat{n}_t^\star = \mathrm{EST}(\widehat{r}_t, (N_\tau, y_\tau)_{\tau=1}^{t-1})$, and $\widehat{p}_t(N) = \widehat{r}_t(N)\widehat{n}_t^\star/|N \cap X|$.
  Choose $N_t \in \arg\max_{N \in \widehat{\mathcal{H}}} \big(\widehat{r}_t(N) + \widehat{p}_t(N)\big)$.
  if $\widehat{r}_t(N_t) + \widehat{p}_t(N_t) + \dfrac{64}{\alpha}\sqrt{\dfrac{d \log(2eT/d) + \log(512T/\delta)}{t-1}} \leq 1$ **then**
    Predict $\emptyset$ for all remaining rounds and stop.
  end if
  Predict N_t and append its recall and precision observations.
end for

APRIL first filters by empirical recall, then queries a hypothesis maximizing empirical recall per item to estimate the target size and filter by estimated precision. In the third stage it updates the estimates using only that stage's nonempty queries. It predicts the empirical maximizer while its upper confidence bound exceeds the empty-set score 1; otherwise it predicts $\emptyset$ for the remaining horizon. This comparison also protects against the case where filtering removes the optimal hypothesis but leaves a nonempty class.

We state the algorithm for $1 \leq d \leq T$ and a nonempty target. All stages stop at the horizon. An empty pool is handled by predicting $\emptyset$; a null recall observation reveals an empty target, after which $\emptyset$ is optimal. If $d = 0$, all hypotheses have the same trace and predicting it gives zero regret. For $T < d$, the bound $\mathrm{Reg}_T \leq T$ suffices. In Algorithm 2, if either filtered class is empty when constructed, predict $\emptyset$ for all remaining rounds and stop. In Stage 3, whenever an estimate is undefined, predict a fixed member of $\widehat{\mathcal{H}}$, append its recall and precision observations, and continue to the next round, skipping the maximization and stopping test.

Theorem 4.3. *Let $\mathcal{H}$ have VC dimension $1 \leq d \leq T$, let the target be nonempty, and let $\alpha \in (0, 1]$, $\delta \in (0, 1)$. In the simultaneous-feedback model, Algorithm 2 satisfies, with probability at least $1 - \delta$,*

$$\mathit{Reg}_T = O\left(\frac{d \log(ed/\alpha^2) + \log(8/\delta)}{\alpha^2} + \frac{1}{\alpha}\sqrt{\big(d \log(2eT/d) + \log(512T/\delta)\big)T} + \alpha T \right).$$

If $\alpha \leq \alpha_{\mathcal{H}}$, the αT term can be omitted. In particular, setting

$$\alpha = \alpha^\star := \min\left\{1, \left(\frac{d \log(eT) + \log(512T/\delta)}{T} \right)^{1/4} \right\}$$

gives

$$\mathit{Reg}_T = O\left(\sqrt{\big(d \log(eT) + \log(512T/\delta)\big)T} + \big(d \log(eT) + \log(512T/\delta)\big)^{1/4} T^{3/4} \right).$$

Remark 4.4. When a positive value $\alpha \leq \alpha_{\mathcal{H}}$ is supplied, the αT term is unnecessary. For fixed positive α, the resulting bound is $\widetilde{O}(\sqrt{dT})$. Without a supplied margin lower bound, the choice $\alpha = \alpha^\star$ gives the general guarantee above. A nonempty filtered class alone does not certify $\alpha \leq \alpha_{\mathcal{H}}$.

To prove Theorem 4.3, we use Lemmas 4.1–4.3 to control the target-size, recall, and precision estimates. The following lemmas assume a nonempty target and $\alpha \in (0, 1]$; filtering statements concern completed first and second stages. Appendix B contains the proofs and establishes survivor quality for arbitrary α; only retention of $N^\star$ requires $\alpha \le \alpha_{\mathcal{H}}$.

Lemma 4.1. *Given a possibly data-dependent estimation function $\widehat{r}(\cdot)$ and any sequence of nonempty sets $N_1, \dots, N_m$ selected before their corresponding fresh precision feedback $(y_t)_{t=1}^m$, where $y_t = \mathbb{1}(v_t \in N^{target})$ with $v_t \sim \mathrm{Unif}(N_t)$, with probability at least $1 - \delta$ the following implication holds: if*

- $|\widehat{r}(N) - r(N)| \le \varepsilon, \forall N \in \{N_t | t \in [m]\}$;
- $\forall N \in \{N_t | t \in [m]\}, p(N) \ge \alpha, r(N) \ge \alpha$ *with* $\alpha \ge 2\varepsilon$,

then the target set size estimator defined in Equation (1) *satisfies*

$$\frac{|\,\mathrm{EST}(\widehat{r}, (N_t, y_t)_{t=1}^m) - |N^{target}||}{|N^{target}|} \le \frac{2}{\alpha}\sqrt{\frac{\log(2/\delta)}{m}} + \frac{2\varepsilon}{\alpha}.$$

In the next lemma, we show that if the hypothesis class is good, i.e., $\alpha_{\mathcal{H}} \ge \alpha$, then with high probability the hypotheses in the filtered $\widehat{\mathcal{H}}$ have non-trivial precision and recall and the optimal hypothesis $N^\star$ is in $\widehat{\mathcal{H}}$.

Lemma 4.2. *For any hypothesis class $\mathcal{H}$ with $\alpha_{\mathcal{H}} \ge \alpha$ and any $X \subseteq \mathcal{X}$, with probability at least $1 - \delta$, we have that for all $N \in \widehat{\mathcal{H}}$, $\min(p(N), r(N)) \ge \frac{3\alpha}{16}$ and $N^\star \in \widehat{\mathcal{H}}$.*

The following lemma bounds the third-stage recall estimates and, once enough observations have been collected, the precision estimates.

Lemma 4.3. *Given a hypothesis class $\mathcal{H}$ of VC dimension $d \ge 1$, fix $\delta \in (0, 1)$ and an integer horizon $T_3 \ge \max\{2, d\}$. Condition on a first-two-stage history with $\widehat{\mathcal{H}} \ne \emptyset$ and $\min(p(N), r(N)) \ge 3\alpha/16$ for every $N \in \widehat{\mathcal{H}}$. With conditional probability at least $1 - \delta$, for every $N \in \widehat{\mathcal{H}}$ and every available third-stage prefix $t \in \{2, \dots, T_3\}$ simultaneously,*

$$|\widehat{r}_t(N) - r(N)| \le \sqrt{\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}}.$$

Whenever $t - 1 \ge \frac{1024}{\alpha^2}\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)$, all EST denominators are positive, and

$$\begin{aligned}|\widehat{p}_t(N) - p(N)| \le\ & \left(1 + \frac{64}{3\alpha}\right)\sqrt{\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}} \\ & + \frac{64}{3\alpha}\,\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}.\end{aligned}$$

5 Discussion

In this paper, we introduced a simple statistical learning framework for identifying an unknown target subset of a domain in the online setting . Under this framework, we showed that a hypothesis class is learnable if and only if it has finite VC dimension. However, a $\log^2 T$ factor gap remains between the upper and lower regret bounds in the realizable setting, which we leave as an open question. In the agnostic setting, we provided an upper bound on regret, but the optimal regret rate remains unknown.

Scope and limitations. Our guarantees concern a fixed target, a known finite pool, and uniform sampling within the predicted and target sets. The motivating applications illustrate possible sources of the two signals; they do not assert that real user behavior is uniform. Our results address statistical learnability and do not establish computational efficiency for arbitrary hypothesis classes. VC dimension characterizes qualitative learnability in our formulation, while the precise minimax regret remains open. The following extensions relax distinct parts of the model.

Unknown available sets. We assume that the finite available set $X \subseteq \mathcal{X}$ is known to the learner. It remains open whether finite VC dimension continues to characterize learnability when X is fixed

but not revealed, so that the learner observes X only through precision and recall feedback from the effective predictions $N_t \cap X$.

Different feedback types and costs. Equal feedback probabilities are a simplifying assumption. If each round independently provides precision feedback with probability $q \in (0, 1)$ and recall feedback with probability $1 - q$, collecting both types takes $O(1/\min\{q, 1-q\})$ rounds in expectation, or $O(\log(T/\delta)/\min\{q, 1-q\})$ rounds per simulated step with high probability. The realizable VC-based algorithm uses only recall feedback, requiring an expected waiting time of $1/(1-q)$ per recall sample, whereas APRIL uses both signals. Recall-only learning has also been studied in the PAC setting (Cohen et al., 2024). Changes to feedback availability are distinct from changing the objective to $g_\lambda = \lambda p + (1-\lambda)r$; Appendix D provides the simulation argument, the realizable weighted-objective extension, and examples illustrating the different informational roles of the two signals. Allowing the learner to choose which signal to request when precision and recall feedback have different costs introduces a trade-off between information and cost. Characterizing this trade-off is an open direction.

More general reward definitions. Infinite available sets require an explicit reference measure, together with suitable sampling and optimization access; a prior on the hypothesis class alone does not define uniform sampling from its sets. For weighted precision–recall objectives, the realizable guarantee extends directly (Appendix D). An agnostic extension requires changing the baseline, margin, filters, and estimated objective together. The optimal dependence on the weights, and nonlinear objectives such as F-measures, remain directions for further study.

Multiple users. This work focuses on learning a single target set. A natural and important extension is to a functional setting, in which the goal is to learn a set-valued function rather than a single set. This richer setting introduces new challenges, such as how to capture shared structure across target sets.

Acknowledgments

Yishay Mansour has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 882396), by the Israel Science Foundation (1357/24), the Blavatnik Foundation at Tel Aviv University and a grant from the Tel Aviv University Center for AI and Data Science (TAD).

Shay Moran acknowledges support from ISF grant 2362/26, by the Technion Center for Machine Learning and Intelligent Systems (MLIS), and by the European Union (ERC, GENERALIZATION, 101039692). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Lee Cohen is supported by the Simons Foundation Collaboration on the Theory of Algorithmic Fairness, and the Simons Foundation investigators award 17351.

References

Agarwal, S., Graepel, T., Herbrich, R., Har-Peled, S., and Roth, D. (2005). Generalization bounds for the area under the ROC curve. J. Mach. Learn. Res., 6:393–425.

Arora, M., Kanjilal, U., and Varshney, D. (2016). Evaluation of information retrieval: precision and recall. International Journal of Indian Culture and Business Management, 12(2):224–236.

Bogatinovski, J., Todorovski, L., Džeroski, S., and Kocev, D. (2022). Comprehensive comparative study of multi-label classification methods. Expert Systems with Applications.

Cohen, L., Mansour, Y., Moran, S., and Shao, H. (2024). Probably approximately precision and recall learning. CoRR.

Cortes, C. and Mohri, M. (2003). AUC optimization vs. error rate minimization. In Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8-13, 2003, Vancouver and Whistler, British Columbia, Canada], pages 313–320. MIT Press.

Cortes, C. and Mohri, M. (2004). Confidence intervals for the area under the ROC curve. In Advances in Neural Information Processing Systems 17 [Neural Information Processing Systems, NIPS 2004, December 13-18, 2004, Vancouver, British Columbia, Canada], pages 305–312.

Elisseeff, A. and Weston, J. (2001). A kernel method for multi-labelled classification. In Advances in Neural Information Processing Systems.

Gao, W. and Zhou, Z.-H. (2011). On the consistency of multi-label learning. In Proceedings of the 24th Annual Conference on Learning Theory, Proceedings of Machine Learning Research.

Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

Han, J., Kamber, M., and Pei, J. (2011). Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.

Joshi, N., Li, G., Bhandari, S., Kasiviswanathan, S. P., Ma, C., and Srebro, N. (2026). Learning to answer from correct demonstrations. In The Fourteenth International Conference on Learning Representations.

Juba, B. and Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 4039–4048.

Kapoor, A., Viswanathan, R., and Jain, P. (2012). Multilabel classification using bayesian compressed sensing. In Advances in Neural Information Processing Systems.

Lattimore, T. and Szepesvári, C. (2020). Bandit Algorithms. Cambridge University Press.

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. Machine learning, 2(4):285–318.

Liu, B., Lee, W. S., Yu, P. S., and Li, X. (2002). Partially supervised classification of text documents. In Proceedings of the Nineteenth International Conference on Machine Learning, pages 387–394.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

Mansouri, F. and Ben-David, S. (2025). Learning from positive and unlabeled examples – finite size sample bounds. In Advances in Neural Information Processing Systems, volume 38.

Mao, A., Mohri, M., and Zhong, Y. (2024). Multi-label learning with stronger consistency guarantees.

McCallum, A. K. (1999). Multi-label text classification with a mixture model trained by em. AAAI'99 workshop on text learning.

Menon, A. K., Rawat, A. S., Reddi, S., and Kumar, S. (2019). Multilabel reductions: what is my loss optimising? In Advances in Neural Information Processing Systems.

Petterson, J. and Caetano, T. (2011). Submodular multi-label learning. In Advances in Neural Information Processing Systems.

Portier, F. (2026). A uniform relative deviation inequality for VC-subgraph classes.

Pour, A. F., Mansouri, F., and Ben-David, S. (2026). Learning with multiple correct answers – regret bounds under different feedback models.

Rosset, S. (2004). Model selection via the AUC. In Brodley, C. E., editor, Machine Learning, Proceedings of the Twenty-first International Conference (ICML 2004), Banff, Alberta, Canada, July 4-8, 2004, volume 69 of ACM International Conference Proceeding Series. ACM.

Sajjadi, M. S., Bachem, O., Lucic, M., Bousquet, O., and Gelly, S. (2018). Assessing generative models via precision and recall. Advances in neural information processing systems, 31.

Schapire, R. E. and Singer, Y. (2000). Boostexter: A boosting-based system for text categorization. Machine Learning.

Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., and Gottschlich, J. (2018). Precision and recall for time series. In Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada, pages 1924–1934.

Torgo, L. and Ribeiro, R. (2009). Precision and recall for regression. In Discovery Science: 12th International Conference, DS 2009, Porto, Portugal, October 3-5, 2009 12, pages 332–346. Springer.

Zhang, M.-L. and Zhou, Z.-H. (2014). A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering.

A Proofs for the Realizable Setting

A.1 Proof of the Online-Classification Reduction

Theorem 3.1. *For any $X \subseteq \mathcal{X}$ and $N^{target} \subseteq X$, for any realizable hypothesis class $\mathcal{H}$, for any online classification algorithm $\mathcal{A}$ with mistake bound M for $\mathcal{H}$, Algorithm 1 will incur regret $\text{Reg}_T \leq M$. By choosing $\mathcal{A}$ to be the standard optimal algorithm by Littlestone (1988), Algorithm 1 will incur regret $\text{Reg}_T \leq \text{Ldim}(\mathcal{H})$, where $\text{Ldim}(\mathcal{H})$ is the Littlestone dimension of $\mathcal{H}$.*

Proof of Theorem 3.1. Since the setting is realizable, $h^* \in \mathcal{H}$ never makes any mistake. Whenever we make a mistake, there are two cases:

- **Precision mistake**: we observe an $v_t \sim \text{Unif}(N_t)$ and $v_t \notin N^{\text{target}}$. Then $\mathcal{A}$ makes a mistake at $(v_t, 0)$.
- **Recall mistake**: we observe an $u_t \sim \text{Unif}(N^{\text{target}})$ and $u_t \notin N_t$. Then $\mathcal{A}$ makes a mistake at $(u_t, 1)$.

Hence, if $\mathcal{A}$ has mistake bound M, we have regret $\text{Reg}_T \leq M$. □

A.2 Proof of the Maximum-Likelihood Guarantee

Theorem 3.2. *For any $X \subseteq \mathcal{X}$ and $N^{target} \subseteq X$, for any realizable hypothesis class $\mathcal{H}$, choosing the maximum likelihood hypothesis in each round achieves regret of $(d\log(T) + \log(T/\delta))\log T$ with probability at least $1-\delta$, where $d = \text{VCdim}(\mathcal{H})$.*

Proof of Theorem 3.2. Let $\ell^{\text{recall}}(N) := 1 - r(N)$ denote the recall loss. We reduce our setting to a standard PAC learning problem: for each $N \in \mathcal{H}$, consider its corresponding binary classifier $h_N(x) = \mathbb{1}(x \in N)$ and we consider PAC learning of $\{h_N | N \in \mathcal{H}\}$. In this PAC learning problem, let the data distribution be uniform over N^{target}. The classification error of h_N under this distribution is

$$\text{err}(h_N) := \mathbb{E}_{x \sim \text{Unif}(N^{\text{target}})}[\mathbb{1}(h_N(x) \neq 1)] = \mathbb{E}_{x \sim \text{Unif}(N^{\text{target}})}[\mathbb{1}(x \notin N)] = \ell^{\text{recall}}(N)\,.$$

Our algorithm can therefore be viewed as ERM in this binary classification problem. By the standard uniform convergence bound for classes with finite VC dimension, for any $t = 1, \ldots, T-1$, with probability at least $1 - \delta/T$, any h_N that is consistent with the t observed samples (i.e., $N \in H_{t+1}$) satisfies

$$\ell^{\text{recall}}(N) = \text{err}(h_N) \leq O\left(\frac{d\log\left(\frac{t}{d}\right) + \log\left(\frac{T}{\delta}\right)}{t}\right).$$

For any $N \in H_{t+1}$, since $N^\star$ is consistent and must belong to H_{t+1}, $N^\star$ has perfect precision and recall, and N_t is minimal, we have $|N_t \cap X| \leq |N^\star \cap X|$ and $\frac{|N^\star \cap X|}{|N^{\text{target}}|} = 1$ The precision of N satisfies

$$\begin{aligned} p(N_t) &= r(N_t) \cdot \frac{|N^{\text{target}}|}{|N_t \cap X|} \\ &\geq r(N_t) \cdot \frac{|N^{\text{target}}|}{|N^\star \cap X|} = r(N_t) \\ &\geq 1 - \frac{d\log\left(\frac{t}{d}\right) + \log\left(\frac{T}{\delta}\right)}{t}\,. \end{aligned}$$

By a union bound over all $t = 1, \ldots, T$, we have that with probability at least $1 - \delta$,

$$\begin{aligned} \text{Reg}_T &\leq \sum_{t=1}^{T} \frac{d\log\left(\frac{t}{d}\right) + \log\left(\frac{T}{\delta}\right)}{t} \\ &\leq \left(d\log\left(\frac{T}{d}\right) + \log\left(\frac{T}{\delta}\right)\right)\log(T). \end{aligned}$$

□

A.3 Proof of the VC Lower Bound

Theorem 3.3 (Lower bound)**.** *For any $d > 0$ and any hypothesis class $\mathcal{H}$ with $\mathrm{VCdim}(\mathcal{H}) = d$, there exists an available set X such that for any (possibly randomized) online learner, there exists a target set N^{target} realizable by $\mathcal{H}$ for which the learner will suffer expected regret $\mathbb{E}[Reg_T] \geq d/50$ for $T \geq d/10$.*

Proof of Theorem 3.3. Let $X = \{x_1, \ldots, x_d\}$ be a shattered set for $\mathcal{H}$. Draw $N^{\text{target}} \subseteq X$ uniformly at random among all $\binom{d}{d/2}$ subsets of size $d/2$. Fix any sequence of predictions $N_1, \ldots, N_{d/10}$. Suppose $t \leq d/10$.

If $|N_t| \geq 2d/3$, at least $N_t - t$ items of N_t that have not been observed yet, and each is in N^{target} w.p. at most $1/2$.

$$\begin{aligned}
\mathbb{E}\left[1 - \text{precision}(N_t)\right] &= 1 - \frac{\mathbb{E}\left[|N_t \cap N^{\text{target}}|\right]}{|N_t|} \\
&\geq 1 - \frac{1}{|N_t|}\left(t + \frac{|N_t| - t}{2}\right) \\
&= 1 - \left(\frac{|N_t|/2}{|N_t|} + \frac{t}{2|N_t|}\right) \\
&\geq \frac{1}{2} - \frac{t}{2|N_t|} \\
&\geq \frac{1}{2} - \frac{\frac{d}{10}}{2 \cdot \frac{2d}{3}} \\
&= \frac{1}{2} - \frac{3}{40} > \frac{1}{3}.
\end{aligned}$$

Where one before the last inequality follows since $|N_t| \geq 2d/3$ and $t \leq d/10$.

Assume $|N_t| < 2d/3$. Since $|N^{\text{target}}| = d/2$,

$$\begin{aligned}
\mathbb{E}\left[1 - \text{recall}(N_t)\right] &= 1 - \mathbb{E}\left[\frac{|N_t \cap N^{\text{target}}|}{|N^{\text{target}}|}\right] \\
&= 1 - \frac{t + \frac{|N_t| - t}{2}}{d/2} \\
&= \frac{d - |N_t| - t}{d} \\
&\geq 1 - \frac{\frac{2d}{3} + \frac{d}{10}}{d} \\
&= \frac{7}{30} > \frac{1}{5}.
\end{aligned}$$

In either case

$$\mathbb{E}\left[\max\{1 - \text{precision}(N_t),\, 1 - \text{recall}(N_t)\}\right] \;\geq\; \frac{1}{5}.$$

Summing over $t = 1, \ldots, T$ gives

$$\mathbb{E}[\text{Reg}_T] \;\geq\; \frac{T}{5} \geq \frac{\lfloor d/10 \rfloor}{5} \geq \frac{d}{50}.$$

Thus, there must exist some choice of N^{target} for which $\text{Reg}_T \geq d/50$, establishing the claimed $\Omega(d)$ lower bound. □

A.4 Proof of Proposition 3.4

Proof of Proposition 3.4. Let $\mathcal{X} = \{x_1, \ldots, x_d\}$ and $\mathcal{H}$ be all subsets of $\mathcal{X}$. In each round $t \leq d$, the algorithm predicts $N_t = \{x_t\}$. After d rounds, we have learned for every $x_t \in X$ whether $x_t \in N^{\text{target}}$, i.e., we learned N^{target}. For future rounds, the algorithm just outputs $N_t = N^{\text{target}}$. □

B Proofs for the Agnostic Setting

B.1 Proof of the APRIL Regret Bound

Theorem 4.3. *Let $\mathcal{H}$ have VC dimension $1 \le d \le T$, let the target be nonempty, and let $\alpha \in (0,1]$, $\delta \in (0,1)$. In the simultaneous-feedback model, Algorithm 2 satisfies, with probability at least $1-\delta$,*

$$\mathit{Reg}_T = O\Bigg(\frac{d\log(ed/\alpha^2)+\log(8/\delta)}{\alpha^2} + \frac{1}{\alpha}\sqrt{\big(d\log(2eT/d)+\log(512T/\delta)\big)T} + \alpha T\Bigg).$$

If $\alpha \le \alpha_{\mathcal{H}}$, the αT term can be omitted. In particular, setting

$$\alpha = \alpha^\star := \min\left\{1, \left(\frac{d\log(eT)+\log(512T/\delta)}{T}\right)^{1/4}\right\}$$

gives

$$\mathit{Reg}_T = O\Bigg(\sqrt{\big(d\log(eT)+\log(512T/\delta)\big)T} + \big(d\log(eT)+\log(512T/\delta)\big)^{1/4}T^{3/4}\Bigg).$$

Proof of Theorem 4.3. For the proof only, abbreviate

$$\eta = \delta/8, \qquad L = d\log(2eT/d)+\log(512T/\delta), \qquad A = \frac{d\log(ed/\alpha^2)+\log(8/\delta)}{\alpha^2}.$$

The first two stages use $T_1+T_2 = O(A)$ rounds. All stages are truncated at the horizon, so if they exhaust it, the trivial bound $\mathrm{Reg}_T \le T$ proves the claim.

The proof of Lemma 4.2 establishes a stronger filtering statement: with failure probability less than η, every survivor has both precision and recall at least $3\alpha/16$, for every $\alpha \in (0,1]$; on the same event, $N^\star$ survives whenever $\alpha \le \alpha_{\mathcal{H}}$. Condition on any first-two-stage history in this event. For a nonempty filtered class, apply Lemma 4.3 with confidence η and a hypothetical third-stage horizon T. Its proof extends the nonempty-query sequence after stopping, so this also covers every actual query prefix. The two events fail with total probability less than $2\eta < \delta$.

Let $s(N) = r(N)+p(N) = 2g(N)$ and

$$G = \max\left\{\frac{1}{2}, \max_{N\in\widehat{\mathcal{H}}} g(N)\right\},$$

where $G = 1/2$ when $\widehat{\mathcal{H}} = \emptyset$. Retention implies $g(N^\star) \le G$ if $\alpha \le \alpha_{\mathcal{H}}$. Otherwise $g(N^\star) = (1+\alpha_{\mathcal{H}})/2$ and $G \ge 1/2$, so

$$g(N^\star) - G \le \frac{\alpha}{2}\,\mathbf{1}\{\alpha > \alpha_{\mathcal{H}}\}. \tag{2}$$

If $\widehat{\mathcal{H}} = \emptyset$, all remaining predictions have reward G, and this inequality suffices.

Suppose $\widehat{\mathcal{H}} \ne \emptyset$. In this proof let $m = t-1$ be the number of preceding nonempty third-stage queries, and put

$$m_0 = \left\lceil \frac{1024L}{\alpha^2} \right\rceil, \qquad B_m = \frac{64}{\alpha}\sqrt{\frac{L}{m}} \quad (m \ge 1).$$

For $1 \le m < m_0$, $B_m > 2$. All defined empirical scores are nonnegative, so the stopping test cannot hold. The fallback also makes a nonempty query whenever $m = 0$ or a denominator is zero. These initial queries contribute at most

$$\min\{T, m_0\} \le \sqrt{Tm_0} = O\left(\frac{\sqrt{LT}}{\alpha}\right)$$

regret relative to G.

For $m \ge m_0$, the sharper estimate proved in (10), with the recall bound, gives

$$\big|(\widehat{r}_t(N)+\widehat{p}_t(N)) - s(N)\big| \le \left(2+\frac{128}{3\alpha}\right)\sqrt{\frac{L}{m}} \le B_m \qquad (N \in \widehat{\mathcal{H}}). \tag{3}$$

Here the logarithm in that concentration result is $d\log(2eT/d) + \log(32T/\eta) \le L$. All EST denominators are now positive. If the stopping test holds, (3) implies $\max_{N\in\widehat{\mathcal{H}}} s(N) \le 1$. Hence $G = 1/2$ and permanently predicting the empty set incurs no further regret relative to G. Otherwise, for the selected empirical maximizer N_t,

$$2G \le \widehat{r}_t(N_t) + \widehat{p}_t(N_t) + B_m, \qquad 2g(N_t) \ge \widehat{r}_t(N_t) + \widehat{p}_t(N_t) - B_m.$$

The first inequality uses both the confidence bound for every survivor and the failed stopping test for the empty-set score. Thus $G - g(N_t) \le B_m$, and the total regret over these queries is at most

$$\frac{64\sqrt{L}}{\alpha}\sum_{m=1}^{T} m^{-1/2} \le \frac{128\sqrt{LT}}{\alpha}.$$

Adding the initial stages and (2) proves the claimed sum bound and its refinement when $\alpha \le \alpha_{\mathcal{H}}$.

For the tuning claim, abbreviate $K_T = d\log(eT)+\log(512T/\delta)$. If $K_T < T$, then $\alpha^\star = (K_T/T)^{1/4}$ and

$$\frac{d}{(\alpha^\star)^2} = d\sqrt{\frac{T}{K_T}} \le \sqrt{dT} \le T.$$

Consequently $d\log(ed/(\alpha^\star)^2) + \log(8/\delta) \le K_T$ and $A \le \sqrt{K_T T}$. Also $L \le K_T + d\log 2 = O(K_T)$, so both $\sqrt{LT}/\alpha^\star$ and $\alpha^\star T$ are $O(K_T^{1/4}T^{3/4})$. If $K_T \ge T$, then $\alpha^\star = 1$ and $\mathrm{Reg}_T \le T \le \sqrt{K_T T}$ suffices. □

B.2 Proof of Lemma 4.1

Lemma 4.1. *Given a possibly data-dependent estimation function $\widehat{r}(\cdot)$ and any sequence of nonempty sets $N_1, \dots, N_m$ selected before their corresponding fresh precision feedback $(y_t)_{t=1}^m$, where $y_t = \mathbb{1}(v_t \in N^{target})$ with $v_t \sim \mathrm{Unif}(N_t)$, with probability at least $1-\delta$ the following implication holds: if*

- $|\widehat{r}(N) - r(N)| \le \varepsilon, \forall N \in \{N_t | t \in [m]\}$;
- $\forall N \in \{N_t | t \in [m]\}, p(N) \ge \alpha, r(N) \ge \alpha$ *with* $\alpha \ge 2\varepsilon$,

then the target set size estimator defined in Equation (1) *satisfies*

$$\frac{|\,\mathrm{EST}(\widehat{r}, (N_t, y_t)_{t=1}^m) - |N^{target}||}{|N^{target}|} \le \frac{2}{\alpha}\sqrt{\frac{\log(2/\delta)}{m}} + \frac{2\varepsilon}{\alpha}.$$

Proof of Lemma 4.1. Write $p_i = p(N_i)$ and $M_m = m^{-1}\sum_{i=1}^m y_i/p_i$. Each N_i is selected before its fresh precision observation, so $y_i/p_i - 1$ has conditional mean zero and conditional range of length $1/p_i \le 1/\alpha$. The martingale version of Hoeffding's inequality therefore gives

$$\Pr\left\{|M_m - 1| > \frac{1}{\alpha}\sqrt{\frac{\log(2/\delta)}{2m}}\right\} \le \delta.$$

If the lower bounds on the selected sets hold only on an event, apply the same inequality to $\mathbf{1}\{p_i \ge \alpha,\ r(N_i) \ge \alpha\}(y_i/p_i - 1)$, defined as zero when the indicator vanishes. Its conditional mean is zero and its conditional range has length at most $1/\alpha$; on histories satisfying the assumptions it is the same sequence. The probability assertion thus applies to the implication in the lemma, without conditioning on a data-dependent estimation event.

Put $\rho = \varepsilon/\alpha \le 1/2$. The recall-error assumption gives $|\widehat{r}(N_i)/r(N_i) - 1| \le \rho$, and all denominators are positive. Using $|N_i \cap X|/|N^{\text{target}}| = r(N_i)/p_i$ and the nonnegativity of y_i/p_i, we obtain the pathwise comparison

$$\frac{M_m}{1+\rho} \le \frac{\mathrm{EST}(\widehat{r}, (N_i, y_i)_{i=1}^m)}{|N^{\text{target}}|} \le \frac{M_m}{1-\rho}.$$

It follows that the relative error is at most

$$\frac{|M_m - 1| + \rho}{1-\rho} \le \frac{\sqrt{2}}{\alpha}\sqrt{\frac{\log(2/\delta)}{m}} + \frac{2\varepsilon}{\alpha} \le \frac{2}{\alpha}\sqrt{\frac{\log(2/\delta)}{m}} + \frac{2\varepsilon}{\alpha}.$$

Only the oracle variables y_i/p_i are used in the martingale argument; the empirical recall function may depend on the entire observed history. □

B.3 Proof of Lemma 4.2

Lemma 4.2. *For any hypothesis class $\mathcal{H}$ with $\alpha_{\mathcal{H}} \geq \alpha$ and any $X \subseteq \mathcal{X}$, with probability at least $1-\delta$, we have that for all $N \in \widehat{\mathcal{H}}$, $\min(p(N), r(N)) \geq \frac{3\alpha}{16}$ and $N^\star \in \widehat{\mathcal{H}}$.*

We use the following uniform relative VC inequality. For an i.i.d. sample of size n from a probability measure P, let P_n be the empirical measure and $\Pi_{\mathcal{A}}$ the growth function of a finite set class $\mathcal{A}$. The two tails in Portier (2026, Theorem 1), whose constants are 1 and 3, give, for every $v > 0$,

$$\Pr\left\{\exists A \in \mathcal{A} : |P_n(A) - P(A)| > v\sqrt{\max\{P_n(A), P(A)\}}\right\} \leq 4\Pi_{\mathcal{A}}(2n)e^{-nv^2/4}. \tag{4}$$

The effective class is finite because X is finite.

Proof of Lemma 4.2. We prove the stronger assertion that every survivor has precision and recall at least $3\alpha/16$ for arbitrary $\alpha \in (0, 1]$; only retention of $N^\star$ requires $\alpha \leq \alpha_{\mathcal{H}}$. For this proof, put

$$\eta = \delta/8, \quad K = d\log\frac{64d}{\alpha^2} + \log\frac{4}{\eta}, \quad n = T_1 = \left\lceil \frac{1024K}{\alpha^2} \right\rceil, \quad Q = d\log\frac{2en}{d} + \log\frac{32}{\eta}.$$

The sample length satisfies $Q \leq 4K$. Indeed, for $u = K/d \geq \log 64$,

$$Q - K \leq d\log\big(2e(16u + 1/64)\big) + \log 8 \leq 3du = 3K.$$

The last inequality follows because $3u - \log(2e(16u + 1/64)) - \log 8$ is increasing and positive for $u \geq \log 64$. This calculation also includes the ceiling in T_1. Using (4), Sauer–Shelah, and $4Q/n \leq \alpha^2/64$ gives, with failure probability at most $\eta/8$,

$$|\widehat{r}_0(N) - r(N)| \leq \alpha/8 \qquad (N \in \mathcal{H}). \tag{5}$$

Thus each $N \in H_1$ satisfies

$$r(N) \geq 3\alpha/4, \qquad \frac{7}{8} \leq \frac{\widehat{r}_0(N)}{r(N)} \leq \frac{7}{6}. \tag{6}$$

For the lower ratio, if $\widehat{r}_0(N) \leq r(N)$, then $r(N) \geq 7\alpha/8$ and $\widehat{r}_0(N)/r(N) \geq \widehat{r}_0(N)/(\widehat{r}_0(N) + \alpha/8) \geq 7/8$. The upper ratio follows from $r(N) \geq 3\alpha/4$.

When $H_1 \neq \emptyset$, denote its selected Stage-2 hypothesis by P within this proof, and set $\overline{y} = T_2^{-1}\sum_{i=1}^{T_2} y_i$ and $p_1 = p(P)$. Conditional on the first stage, Hoeffding gives

$$\Pr\{|\overline{y} - p_1| > \alpha/8\} \leq 2e^{-T_2\alpha^2/32} \leq 2\eta^2. \tag{7}$$

Pilot maximization implies

$$\widehat{p}_0(N) = \overline{y}\,\frac{\widehat{r}_0(N)/|N \cap X|}{\widehat{r}_0(P)/|P \cap X|} \leq \overline{y} = \widehat{p}_0(P) \qquad (N \in H_1).$$

If $\widehat{\mathcal{H}} \neq \emptyset$, then $\overline{y} \geq 7\alpha/16$, whence $p_1 \geq 5\alpha/16$ and $\overline{y}/p_1 \leq 7/5$ on (7)'s good event. For each survivor, (6) now gives

$$\frac{\widehat{p}_0(N)}{p(N)} = \frac{\widehat{r}_0(N)}{r(N)}\,\frac{r(P)}{\widehat{r}_0(P)}\,\frac{\overline{y}}{p_1} \leq \frac{7}{6}\frac{8}{7}\frac{7}{5} = \frac{28}{15}.$$

Consequently

$$p(N) \geq \frac{15\alpha}{64} > \frac{3\alpha}{16}, \qquad r(N) \geq \frac{3\alpha}{4} \qquad (N \in \widehat{\mathcal{H}}). \tag{8}$$

These assertions are vacuous if either filtered class is empty.

If $\alpha \leq \alpha_{\mathcal{H}}$, then $p(N^\star), r(N^\star) \geq \alpha$, and (5) implies $\widehat{r}_0(N^\star) \geq 7\alpha/8$ and $\widehat{r}_0(N^\star)/r(N^\star) \geq 7/8$. Thus $N^\star \in H_1$, and pilot maximization gives

$$p_1 \geq \frac{6}{7}\,|N^{\text{target}}|\frac{\widehat{r}_0(P)}{|P \cap X|} \geq \frac{6}{7}\,|N^{\text{target}}|\frac{\widehat{r}_0(N^\star)}{|N^\star \cap X|} \geq \frac{3}{4}p(N^\star) \geq \frac{3\alpha}{4}.$$

Therefore $\overline{y}/p_1 \geq 5/6$ and

$$\widehat{p}_0(N^\star) \geq p(N^\star)\frac{7}{8}\frac{6}{7}\frac{5}{6} \geq \frac{5\alpha}{8} > \frac{7\alpha}{16},$$

which proves $N^\star \in \widehat{\mathcal{H}}$. The total failure probability is at most $\eta/8 + 2\eta^2 < \eta < \delta$. This also proves the stronger arbitrary-α survivor assertion used in the regret proof. □

B.4 Proof of Lemma 4.3

Lemma 4.3. *Given a hypothesis class $\mathcal{H}$ of VC dimension $d \geq 1$, fix $\delta \in (0,1)$ and an integer horizon $T_3 \geq \max\{2, d\}$. Condition on a first-two-stage history with $\widehat{\mathcal{H}} \neq \emptyset$ and $\min(p(N), r(N)) \geq 3\alpha/16$ for every $N \in \widehat{\mathcal{H}}$. With conditional probability at least $1-\delta$, for every $N \in \widehat{\mathcal{H}}$ and every available third-stage prefix $t \in \{2, \dots, T_3\}$ simultaneously,*

$$|\widehat{r}_t(N) - r(N)| \leq \sqrt{\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}}.$$

Whenever $t - 1 \geq \frac{1024}{\alpha^2}\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)$, all EST denominators are positive, and

$$|\widehat{p}_t(N) - p(N)| \leq \left(1 + \frac{64}{3\alpha}\right)\sqrt{\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}} + \frac{64}{3\alpha}\,\frac{4\big(d\log(2eT_3/d) + \log(32T_3/\delta)\big)}{t-1}.$$

Proof of Lemma 4.3. Fix a first-two-stage history as in the statement. For analysis, extend the nonempty-query sequence after stopping by fresh queries of a fixed survivor. Each actual history is a prefix of this adapted sequence. The recall observations remain i.i.d. uniform on $N^{\text{target}} \cap X$, irrespective of the selected sets. Within this proof, abbreviate

$$b = 3\alpha/16, \qquad L_3 = d\log(2eT_3/d) + \log(32T_3/\delta), \qquad \varepsilon_t = 2\sqrt{\frac{L_3}{t-1}}.$$

Since $d \leq T_3$, the growth function at every sample size used here is bounded by $(2eT_3/d)^d$. Apply (4) at each prefix and take a union bound. Because both probabilities in its maximum are at most one, the recall bounds

$$|\widehat{r}_t(N) - r(N)| \leq \varepsilon_t \qquad (N \in \widehat{\mathcal{H}},\ 2 \leq t \leq T_3) \tag{9}$$

hold with failure probability at most $4T_3(2eT_3/d)^d e^{-L_3} = \delta/8$.

For the oracle average $M_{t-1} = (t-1)^{-1}\sum_{i=1}^{t-1} y_i/p(N_i)$, the martingale Hoeffding argument from the proof of Lemma 4.1, followed by a union bound, gives

$$|M_{t-1} - 1| \leq \frac{1}{b}\sqrt{\frac{\log(4T_3/\delta)}{2(t-1)}} \leq \frac{\varepsilon_t}{2b} \qquad (2 \leq t \leq T_3),$$

with additional failure probability at most $\delta/2$. The two events thus hold jointly with probability at least $1-\delta$.

Under the displayed lower bound on $t-1$ in the lemma, $\varepsilon_t \leq \alpha/16 = b/3$. Put $\rho_t = \varepsilon_t/b \leq 1/3$. Every survivor has recall at least b, so all empirical recall denominators are positive and $|\widehat{r}_t(N)/r(N) - 1| \leq \rho_t$. The same nonnegative sandwich as in the preceding estimator proof yields

$$\frac{M_{t-1}}{1+\rho_t} \leq \frac{\widehat{n}_t^\star}{|N^{\text{target}}|} \leq \frac{M_{t-1}}{1-\rho_t}.$$

Multiplying by $\widehat{r}_t(N)/r(N)$ bounds $\widehat{p}_t(N)/p(N)$ between $M_{t-1}(1-\rho_t)/(1+\rho_t)$ and $M_{t-1}(1+\rho_t)/(1-\rho_t)$. Since $p(N) \leq 1$, this implies

$$\begin{aligned}|\widehat{p}_t(N) - p(N)| &\leq \frac{2\rho_t + (1+\rho_t)|M_{t-1} - 1|}{1-\rho_t} \\ &\leq \frac{2\rho_t + (4/3)(\rho_t/2)}{2/3} = 4\rho_t = \frac{64}{3\alpha}\varepsilon_t.\end{aligned} \tag{10}$$

This is stronger than the stated bound $(1 + 64/(3\alpha))\varepsilon_t + (64/(3\alpha))\varepsilon_t^2$, which retains the original two-term form. The argument is conditional on the initial filtering history and uniform over all later prefixes, so it remains valid at the algorithm's stopping decision. □

B.5 Proof of the Proper-Learning Lower Bound

Theorem 4.1. *There exists a hypothesis class $\mathcal{H}$ with $\mathrm{VCdim}(\mathcal{H}) = 1$ such that, for any T, there exists an available set X and target set $N^{target} \subseteq X$ for which any proper learning algorithm incurs expected regret $\mathbb{E}[Reg_T] \geq C \cdot T$ for some universal numerical constant $C > 0$.*

Proof of Theorem 4.1. We will construct an example and reduce the problem to a bandit problem. Consider a set of k non-overlapping hypotheses $\mathcal{H} = \{N_1, \ldots, N_k\}$. The size $|N|$ of each hypothesis N is identical and is very huge, i.e., $|N| = n \gg T$ for all $N \in \mathcal{H}$. Let available set $X = \mathcal{X} = \cup_{N \in \mathcal{H}} N$.

Now we construct two worlds with different random generation ways of the target set.

In world I, the target set N^{target} is created randomly by selecting $(\frac{1}{2} + \varepsilon)n$ items from the hypothesis N_1 and $(\frac{1}{2} - \frac{\varepsilon}{k-1})n$ from each of the remaining $k-1$ hypotheses, $N_2, \ldots, N_k$. Then no matter what N^{target} is generated, $|N^{\text{target}}| = \frac{kn}{2}$ and the recalls and precisions of hypotheses are

$$r(N_1) = \frac{1 + 2\varepsilon}{k}, p(N_1) = 1/2 + \varepsilon,$$
$$r(N_i) = \frac{1 - \frac{2\varepsilon}{k-1}}{k}, p(N_i) = 1/2 - \frac{\varepsilon}{k-1}, \forall i > 1.$$

For any fixed policy π, let $\mathcal{P}_{\text{I}}$ denote the distribution over the sequence of decisions and feedbacks $((N_1, u_1, v_1, y_1), \ldots, (N_T, u_T, v_T, y_T))$ in world I.

Under this fixed policy π, let T_i denote the number of rounds we select hypothesis N_i. Pick $i' = \arg\min_{i>1} \mathbb{E}_{\mathcal{P}_{\text{I}}}[T_i]$. Hence $\mathbb{E}_{\mathcal{P}_{\text{I}}}[T_i] \leq \frac{T}{k-1}$. Then we construct world II as follows. In world II, the target set N^{target} is created randomly by selecting $(\frac{1}{2} + \varepsilon)n$ items from the hypothesis N_1, selecting $(\frac{1}{2} + 2\varepsilon)n$ items from $N_{i'}$, and $(\frac{1}{2} - \frac{3\varepsilon}{k-2})n$ items from each of the remaining hypotheses. Then no matter what N^{target} is generated, $|N^{\text{target}}| = \frac{kn}{2}$ and the recalls and precisions of hypotheses are

$$r(N_1) = \frac{1 + 2\varepsilon}{k}, p(N_1) = 1/2 + \varepsilon,$$
$$r(N_{i'}) = \frac{1 + 4\varepsilon}{k}, p(N_{i'}) = 1/2 + 2\varepsilon,$$
$$r(N_i) = \frac{1 - \frac{6\varepsilon}{k-2}}{k}, p(N_i) = 1/2 - \frac{3\varepsilon}{k-2}, \forall i \neq 1, i'.$$

Let $\mathcal{P}_{\text{II}}$ denote the distribution over the sequence of decisions and feedbacks over T steps $((N_1, u_1, v_1, y_1), \ldots, (N_T, u_T, v_T, y_T))$ in world II. Following the standard proof techniques in minimax lower bounds for bandit problems, we want to show that the policy will incur high regret in at least one of the world. To disentangle with the choice of N^{target}, instead of considering $\mathcal{P}_{\text{I}}$ and $\mathcal{P}_{\text{II}}$ themselves directly, we consider their alternatives.

At each round t, let i_t denote index of the hypothesis N_i in which the recall feedback $u_t \sim \mathrm{Unif}(N^{\text{target}})$ falls into, i.e., $u_t \in N_{i_t}$. Notice that the probability of $i_t = i$ is exactly $r(N_i)$. The generation process of $\mathcal{P}_{\text{I}}$ is equivalent to:

- Sample N^{target} by randomly picking $r(N_i) \cdot n$ items from N_i for all $i = 1, \ldots, k$.
- Sample i_t from a categorical distribution $\mathcal{C} = (r(N_1), r(N_2), \ldots, r(N_k))$.
- Sample $u_t \sim \mathrm{Unif}(N_{i_t} \cap N^{\text{target}})$.
- For any $N_t = N$ chosen by policy π at round t, sample $v_t \sim \mathrm{Unif}(N)$ and $y_t = \mathbb{1}(v_t \in N^{\text{target}})$.

Now we construct an alternative $\mathcal{P}'_{\text{I}}$ for $\mathcal{P}_{\text{I}}$:

- Sample i_t from the same categorical distribution $\mathcal{C} = (r(N_1), r(N_2), \ldots, r(N_k))$.

- Sample $u_t \sim \mathrm{Unif}(N_{i_t})$.
- For any $N_t = N$ chosen by policy π at round t, again $v_t \sim \mathrm{Unif}(N)$ is still sampled uniformly at random from N. But we choose $y_t \sim \mathrm{Ber}(p(N))$.

We apply the same procedure of generating an alternative $\mathcal{P}'_{\text{II}}$ for $\mathcal{P}_{\text{II}}$. We prove that $\mathcal{P}$ and $\mathcal{P}'$ are close enough, i.e., $\mathrm{TV}(\mathcal{P}'_{\text{I}}, \mathcal{P}_{\text{I}}) \to 0$ and $\mathrm{TV}(\mathcal{P}'_{\text{II}}, \mathcal{P}_{\text{II}}) \to 0$ when $n \to \infty$ by Lemma B.1.

Now we are ready to prove the lower bound based on the standard proof techniques for bandit minimax lower bounds, see Chapter 15.2 of Lattimore and Szepesvári (2020) for example. In world I, no matter what N^{target} is generated, the best hypothesis is N_1 and the regret is at least $\frac{\varepsilon T}{4}$ if $T_1 \le \frac{T}{2}$. However, we will suffer regret of $\frac{\varepsilon T}{4}$ if $T_1 > \frac{T}{2}$ in world II. Let $\mathrm{Reg}_T(\mathcal{P}_{\text{I}})$ and $\mathrm{Reg}_T(\mathcal{P}_{\text{II}})$ denote expected regret in worlds I and II respectively. Then we have

$$
\begin{aligned}
&\mathrm{Reg}_T(\mathcal{P}_{\text{I}}) + \mathrm{Reg}_T(\mathcal{P}_{\text{II}}) \\
>&\mathbb{P}_{\mathcal{P}_{\text{I}}}\left(T_1 \le \frac{T}{2}\right)\frac{\varepsilon T}{4} + \mathbb{P}_{\mathcal{P}_{\text{II}}}\left(T_1 > \frac{T}{2}\right)\frac{\varepsilon T}{4} \\
\ge&\mathbb{P}_{\mathcal{P}'_{\text{I}}}\left(T_1 \le \frac{T}{2}\right)\frac{\varepsilon T}{4} + \mathbb{P}_{\mathcal{P}'_{\text{II}}}\left(T_1 > \frac{T}{2}\right)\frac{\varepsilon T}{4} - \mathrm{TV}(\mathcal{P}'_{\text{I}}, \mathcal{P}_{\text{I}}) - \mathrm{TV}(\mathcal{P}'_{\text{II}}, \mathcal{P}_{\text{II}}) \\
\ge&\frac{\varepsilon T}{8}\exp(-D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}} \| \mathcal{P}'_{\text{II}}))\,,
\end{aligned}
$$

where the last inequality applies the Bretagnolle-Huber inequality (Theorem 14.2 of Lattimore and Szepesvári (2020)) and Lemma B.1. It remains to upper bound $D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}} \| \mathcal{P}'_{\text{II}})$. By chain rule of KL divergence and Lemma 15.1 of Lattimore and Szepesvári (2020), we have

$$
D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}} \| \mathcal{P}'_{\text{II}}) = D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}}(i_{1:T}) \| \mathcal{P}'_{\text{II}}(i_{1:T})) + \sum_i \mathbb{E}_{\mathcal{P}'_{\text{I}}}[T_i]\, D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}}(v, y|N_i) \| \mathcal{P}'_{\text{II}}(v, y|N_i))\,,
$$

where $\mathcal{P}(i_{1:T})$ denotes the distribution of sequence $i_{1:T}$ and $\mathcal{P}(v, y|N_i)$ denotes the distribution of precision feedback (v, y) when selecting N_i under generation process $\mathcal{P}$.

By Lemma B.2, we have

$$
D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}}(i_{1:T}) \| \mathcal{P}'_{\text{II}}(i_{1:T})) \le T \cdot \frac{8\varepsilon^2}{k}\,.
$$

As aforementioned, the distribution over v is identical in both worlds, but y is sampled from $\mathrm{Ber}(p(N_i))$. For any two Bernoulli distributions parametrized by q, q', we have

$$
D_{\mathrm{KL}}(\mathrm{Ber}(q) \| \mathrm{Ber}(q')) \le \frac{(q - q')^2}{q'(1 - q')}\,.
$$

Hence, we have

$$
\sum_i \mathbb{E}_I[T_i]\, D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}}(v, y|N_i) \| \mathcal{P}'_{\text{II}}(v, y|N_i)) \lesssim T \cdot \frac{16\varepsilon^2}{k^2} + \frac{T}{k-1} \cdot 16\varepsilon^2\,.
$$

Hence, we have

$$
D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}} \| \mathcal{P}'_{\text{II}}) \le T \cdot \frac{40\varepsilon^2}{k}
$$

Then by setting $\varepsilon = \min\{\sqrt{\frac{k}{80T}}, 1\}$, we have $D_{\mathrm{KL}}(\mathcal{P}'_{\text{I}} \| \mathcal{P}'_{\text{II}}) \le \frac{1}{2}$ when $k \le T$. Thus,

$$
\mathrm{Reg}_T(\mathcal{P}_{\text{I}}) + \mathrm{Reg}_T(\mathcal{P}_{\text{II}}) \ge \frac{\varepsilon T}{8}\exp(-\frac{1}{2}) \ge C\sqrt{kT},
$$

for a universal constant $C > 0$. By setting $k = T$, we are done with the proof. □

Lemma B.1. *TV*$(\mathcal{P}'_I, \mathcal{P}_I) \to 0$ *and TV*$(\mathcal{P}'_{II}, \mathcal{P}_{II}) \to 0$ *when* $n \to \infty$.

Proof of Lemma B.1. When $n \to \infty$, it is almost sure that there is no reappearance in $(u_1, u_2, \ldots, u_T)$. Conditional on no reappearance, every sequence will appear with equal probability under both $\mathcal{P}_{\mathrm{I}}$ and $\mathcal{P}'_{\mathrm{I}}$. Hence $\mathrm{TV}(\mathcal{P}'_{\mathrm{I}}(u_{1:T}), \mathcal{P}_{\mathrm{I}}(u_{1:T})) \to 0$.

For any $t = 1, \ldots, T$, let $F_t = ((N_1, u_1, v_1, y_1), \ldots, (N_{t-1}, u_{t-1}, v_{t-1}, y_{t-1}), N_t)$ denote the history up to round t. In round t, no matter what $N_t = N$ is chosen, the distribution of $v_t \sim \mathrm{Unif}(N)$ is identical in $\mathcal{P}'_{\mathrm{I}}$ and $\mathcal{P}_{\mathrm{I}}$. Since $|N| = n \to \infty$ is large for any $N \in \mathcal{H}$, it is almost sure that there is no reappearance in $(v_1, v_2, \ldots, v_T)$. Conditional on no reappearance,

$$\mathcal{P}_{\mathrm{I}}(y_t = 1 | F_t) = \mathbb{E}\left[\mathbb{P}\left[v_t \in N^{\text{target}} \,\middle|\, N^{\text{target}}, F_t\right]\right] = \mathbb{E}\left[\frac{|N \cap N^{\text{target}}| - t + 1}{n}\right] \to p(N)$$

as $n \to \infty$. □

Lemma B.2. *We have* $\frac{1}{T} D_{\mathrm{KL}}(\mathcal{P}'_I(i_{1:T}) \| \mathcal{P}'_{II}(i_{1:T})) \le \frac{20\varepsilon^2}{k}$ *for all* $\varepsilon \le \frac{1}{8}$ *and* $k \ge 11$.

Proof of Lemma B.2. In world $\mathcal{P}'_{\mathrm{I}}$, $i_{1:T}$ are i.i.d samples from $\mathcal{P}'_{\mathrm{I}}(i)$ with $\mathbb{P}(i_t = 1) = \frac{1+2\varepsilon}{k}$ and $\mathbb{P}(i_t = i) = \frac{1 - \frac{2\varepsilon}{k-2}}{k}, \forall i \notin \{1, i'\}$. In world II, $i_{1:T}$ are i.i.d samples from $\mathcal{P}'_{\mathrm{II}}(i)$, with $\mathbb{P}(i_t = 1) = \frac{1+2\varepsilon}{k}$, $\mathbb{P}(i_t = i') = \frac{1+4\varepsilon}{k}$, and $\mathbb{P}(i_t = i) = \frac{1 - \frac{6\varepsilon}{k-2}}{k}, \forall i \notin \{1, i'\}$.

We compute the KL divergence from $\mathcal{P}'_{\mathrm{I}}(i)$ to $\mathcal{P}'_{\mathrm{II}}(i)$:

$$\begin{aligned} D_{\mathrm{KL}}(\mathcal{P}'_{\mathrm{I}}(i) \| \mathcal{P}'_{\mathrm{II}}(i)) =& \frac{1+a}{k} \log\left(\frac{1+a}{1+4\varepsilon}\right) + (k-2) \cdot \frac{1+a}{k} \log\left(\frac{1+a}{1+c}\right) \\ =& \frac{1+a}{k} \left((k-1)\log(1+a) - \log(1+4\varepsilon) - (k-2)\log(1+c)\right), \end{aligned}$$

where $a := -\dfrac{2\varepsilon}{k-1}$ and $c := -\dfrac{6\varepsilon}{k-2}$.

For any $x \in [-\frac{1}{2}, 1]$, we have

$$\ln(1+x) \le x - \frac{x^2}{2} + \frac{x^3}{3},$$

and

$$\ln(1+x) \ge x - x^2.$$

Then we have

$$\begin{aligned} D_{\mathrm{KL}}(\mathcal{P}'_{\mathrm{I}}(i) \| \mathcal{P}'_{\mathrm{II}}(i)) =& \frac{1+a}{k} \left((k-1)(a - \frac{1}{2}a^2 + \frac{a^3}{3}) - (4\varepsilon - 16\varepsilon^2) - (k-2)(c - c^2)\right) \\ \le& \frac{1+a}{k} \left(16\varepsilon^2 + (k-2)c^2\right) \\ \le& \frac{20\varepsilon^2}{k}, \end{aligned}$$

when $k \ge 11$. □

C Finite Available Pools and the Ambient Domain

Our upper bounds also apply when $X = \mathcal{X}$ is finite. However, our learnability definition requires a single sublinear regret bound uniformly over all finite available sets. Lower bounds for this uniform problem need not hold on every fixed pool: on any one finite X, the restricted class $\{N \cap X : N \in \mathcal{H}\}$ is finite, so finite-action bandit learning gives sublinear regret as $T \to \infty$, with a bound that may depend on $|X|$. In particular, the lower bounds must not be interpreted as applying to every fixed finite pool. An infinite available set is outside the present model; defining feedback there would require an explicit sampling measure.

To clarify why the distinction matters in that case, consider the following example in the realizable setting:

- Let the domain $\mathcal{X}$ be the disjoint union of $D \cup E_1 \cup \cdots \cup E_{2^d}$, where $|D| = d$ and each $|E_i|$ is much larger than d.
- There is a one-to-one mapping between each subset $C \subseteq D$ (there are 2^d of them) and a corresponding E_i.
- Define the hypothesis class:

$$\mathcal{H} = \{C \cup E_i : C \subseteq D \text{ and } E_i \text{ corresponds to } C\}.$$

  This class has a VC dimension d.

Take the sets E_i to be finite, with $|E_i| \geq d$, and let $X = \mathcal{X}$ and $C \cup E_i$ be the target. A recall example lands in E_i with probability $|E_i|/(|C| + |E_i|) \geq 1/2$. Such an example identifies i and hence the target exactly. The expected number of recall observations until this occurs is at most 2, so a learner that predicts the identified target thereafter has constant expected regret, independently of d (also under the original equal-probability feedback protocol). Thus VC dimension alone does not lower-bound regret on this particular pool. In contrast, restricting to $X = D$ produces the full powerset class 2^D, to which the worst-case lower bound of Theorem 3.3 applies.

D Asymmetric Feedback and the Roles of Precision and Recall

Feedback frequencies and the evaluation objective. Let each round provide precision feedback with probability $q \in (0, 1)$ and recall feedback with probability $1 - q$, independently of the history. For this paragraph, keep the evaluation objective $g = (p + r)/2$ unchanged. The first precision and first recall observations can be collected by holding the same prediction until both have arrived. The expected waiting time is

$$\frac{1}{q} + \frac{1}{1-q} - 1 = O\left(\frac{1}{\min\{q, 1-q\}}\right).$$

This is a sampling overhead in expectation. For a high-probability simulation within an actual horizon T, assuming q is known, set

$$m = \left\lceil \frac{\log(4T/\delta)}{\min\{q, 1-q\}} \right\rceil$$

and repeat each virtual prediction for a block of m actual rounds, supplying the paired-feedback learner with the first observation of each type. In any block, the probability of missing a type is at most $q^m + (1-q)^m \leq 2e^{-m \min\{q,1-q\}}$. A union bound over at most T blocks makes the simulation failure probability at most $\delta/2$. For $B = \lfloor T/m \rfloor \geq 1$, a paired-feedback regret guarantee $R(B, \delta/2)$ therefore transfers to

$$\mathrm{Reg}_T \leq mR(B, \delta/2) + m$$

with probability at least $1 - \delta$; the last term covers an unfinished block. If $B = 0$, the trivial bound $\mathrm{Reg}_T \leq T$ applies. The realizable maximum-likelihood learner only waits for recall observations, whose expected inter-arrival time is $1/(1-q)$. The endpoints $q = 0$ and $q = 1$ are not covered by the paired-feedback simulation.

Feedback frequencies and reward weights are separate modeling choices. If one instead retains the convention that the realized reward is the indicator from the observed feedback type, its conditional mean becomes $qp + (1-q)r$, rather than $(p + r)/2$.

Weighted objectives in the realizable setting. Consider $g_\lambda(N) = \lambda p(N) + (1-\lambda)r(N)$ for $\lambda \in [0, 1]$. The maximum-likelihood rule chooses a smallest positive-consistent effective set, so in the realizable case $|N_t| \leq |N^{\text{target}}|$. For nonempty target and prediction, this implies $p(N_t) \geq r(N_t)$; the empty-set conventions give the same inequality in the remaining cases. Consequently,

$$1 - g_\lambda(N_t) \leq 1 - r(N_t) \leq 2\big(1 - g(N_t)\big),$$

Thus any realizable regret bound for this rule transfers to every fixed λ with a factor of at most two, under the same feedback protocol. In particular, the stated $O(d \log^2 T)$ rate transfers up to a constant, retaining any sampling overhead from unequal feedback frequencies. At $\lambda = 0$, the full pool achieves perfect recall; at $\lambda = 1$, the empty prediction achieves perfect precision under our convention.

For the agnostic setting, reweighting estimated precision and recall gives

$$|\widehat{g}_{\lambda,t}(N) - g_\lambda(N)| \leq \lambda|\widehat{p}_t(N) - p(N)| + (1-\lambda)|\widehat{r}_t(N) - r(N)|.$$

This transfers uniform estimation accuracy on a retained candidate class, but is not by itself a complete regret argument: the algorithm must also retain a suitable comparator or fall back to an adequate baseline. The trivial predictions guarantee at least $b_\lambda = \max\{\lambda, 1-\lambda\}$; one may use $\emptyset$ when $\lambda \geq 1/2$ and X otherwise. Thus the baseline, optimality margin, and filtering conditions must be adapted along with the objective. We do not state a separate agnostic weighted-regret theorem here.

The value of recall feedback. Let X contain n items, let $\mathcal{H}$ be the class of all singleton subsets of X, and let the target be one unknown singleton. One recall observation identifies the target. A proper learner with only precision feedback tests one singleton at a time and needs $\Omega(n)$ tests in the worst case. This comparison concerns proper precision-only learners; it does not assume that arbitrary subset queries are restricted to singleton tests.

The value of precision feedback. For an agnostic proper-learning example, take $\mathcal{H} = \{\{a\}, \{b\}\}$ and a set S disjoint from $\{a, b\}$. The target is either $S \cup \{a\}$ or $S \cup \{b\}$. Recall observations distinguish these targets only upon observing a or b, which takes $|S| + 1$ observations in expectation. Under the uniform prior over the two targets, before this happens every proper prediction has probability $1/2$ of selecting the wrong singleton, with a constant reward gap. Thus for $T = \Theta(|S|)$, every proper recall-only learner has $\Omega(|S|)$ expected regret on at least one target. One precision observation from either singleton distinguishes the targets. This example does not establish a necessity claim for improper learners.

In the realizable case, the online-classification reduction can also provide a smaller bound than the general VC-based guarantee: for a class of n distinct one-dimensional thresholds, its $O(\log n)$ mistake bound can improve on $O(\log^2 T)$ when $\log n \ll \log^2 T$. This comparison of upper bounds does not prove that precision feedback is necessary for the improvement.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our abstract and introduction provide a summary of contributions that accurately reflect the new model and the results of this paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See discussion.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provided all the assumptions and complete proofs (some are in the appendix).

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: The paper does not include experiments. All results are carefully proven and do not require experiments to establish correctness.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: This is a theoretical paper. All results are carefully proven and do not require experiments to establish correctness.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: This is a theoretical paper. All results are carefully proven and do not require experiments to establish correctness.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: This is a theoretical paper. All results are carefully proven and do not require experiments to establish correctness.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: This is a theoretical paper. All results are carefully proven and do not require experiments to establish correctness.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have carefully read and comply with the NeurIPS code of ethics. This work is theoretical and we do not foresee harmful consequences.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: We note that our work is theoretical. We do not foresee any societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The work is entirely theoretical; we do not release new data or models.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: We do not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: We only used LLMs for editing.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.