Sample-Efficient Agnostic Boosting

Udaya Ghai
Amazon
ughai@amazon.com

Karan Singh
Tepper School of Business
Carnegie Mellon University
karansingh@cmu.edu

Abstract

The theory of boosting provides a computational framework for aggregating approximate weak learning algorithms, which perform marginally better than a random predictor, into an accurate strong learner. In the realizable case, the success of the boosting approach is underscored by a remarkable fact that the resultant sample complexity matches that of a computationally demanding alternative, namely Empirical Risk Minimization (ERM). This in particular implies that the realizable boosting methodology has the potential to offer computational relief without compromising on sample efficiency.

Despite recent progress, in agnostic boosting, where assumptions on the conditional distribution of labels given feature descriptions are absent, ERM outstrips the agnostic boosting methodology in being quadratically more sample efficient than all known agnostic boosting algorithms. In this paper, we make progress on closing this gap, and give a substantially more sample efficient agnostic boosting algorithm than those known, without compromising on the computational (or oracle) complexity. A key feature of our algorithm is that it leverages the ability to reuse samples across multiple rounds of boosting, while guaranteeing a generalization error strictly better than those obtained by blackbox applications of uniform convergence arguments. We also apply our approach to other previously studied learning problems, including boosting for reinforcement learning, and demonstrate improved results.

1 Introduction

A striking observation in statistical learning is that given a small number of samples it is possible to learn the best classifier from an almost exponentially large class of predictors. In fact, it is possible to do using a conceptually straightforward procedure – Empirical Risk Minimization (ERM) – that finds a classifier that is maximally consistent with the collected samples. Substantiating this observation, the fundamental theorem of statistical learning (e.g., [SSBD14]) states that with high probability ERM can guarantee $\varepsilon-$excess population error with respect to the best classifier from a finite, but large hypothesis class $\mathcal{H}$ given merely $m_{\text{agnostic}} \approx (\log|\mathcal{H}|)/\varepsilon^2$ identically distributed and independent (IID) samples of pairs of features and labels from the population distribution. Under an additional assumption – that of *realizability* – guaranteeing that there exists a perfect classifier in the hypothesis class achieving zero error, a yet quadratically smaller number $m_{\text{realizable}} \approx (\log|\mathcal{H}|)/\varepsilon$ of samples suffice. In the absence of such assumption, the learning problem is said to take place in the *agnostic* setting, i.e., under of a lack of belief in the ability of any hypothesis to perfectly fit the observed data.

This ability to generalize to the population distribution and successfully (PAC) learn from an almost exponentially large, and hence expressive, hypothesis class given limited number of examples suggests that the primary bottleneck for efficient learning is computational. Indeed, even with modest sample requirements, finding a maximally consistent hypothesis within an almost exponentially large class via, say, enumeration or global search, is generally computationally intractable.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

https://doi.org/10.52202/079017-3190

	Sample Complexity	Oracle Complexity
[KK09]	$(\log \lvert\mathcal{B}\rvert)/\gamma^4\varepsilon^4$	$1/\gamma^2\varepsilon^2$
[BCHM20]	$(\log \lvert\mathcal{B}\rvert)/\gamma^4\varepsilon^6$	$1/\gamma^2\varepsilon^2$
Theorem 4	$(\log \lvert\mathcal{B}\rvert)/\gamma^3\varepsilon^3$	$(\log \lvert\mathcal{B}\rvert)/\gamma^2\varepsilon^2$
Theorem 9 (in Appendix B)	$(\log \lvert\mathcal{B}\rvert)/\gamma^3\varepsilon^3 + (\log \lvert\mathcal{B}\rvert)^3/\gamma^2\varepsilon^2$	$1/\gamma^2\varepsilon^2$
ERM (no boosting)	$(\log \lvert\mathcal{B}\rvert)/\gamma^2\varepsilon^2$	$+\infty$ (inefficient)

Table 1: A comparison between sample and oracle complexities (i.e., number of weak learning calls) of the present results and previous works, in each case to achieve ε-excess population error. Here we suppress polylogarithmic factors. We make progress on closing the sample complexity gap between ERM, which is computationally inefficient, and boosting-based approaches. The γ-weak leaner outputs a hypothesis from the base class $\mathcal{B}$, which is usually substantially smaller than $\mathcal{H}$ against which the final agnostic learning guarantee holds. In practice, boosting is used with learners with small values of $\log |\mathcal{B}|$. See Definition 1 for details. See the paragraph following Theorem 1 in [KK09] and Section 3.3 in [BCHM20] for derivation of these bounds. See also Theorem 2.14 in [AGHM21] for a bound on the expressivity of the boosted class to derive ERM's sample complexity.

It is against this backdrop that the theory of boosting offers a compelling alternative. The starting point is the realization that often, both in practice and in theory, it is easy to construct simple, yet inaccurate *rules-of-thumbs* (or *weak learners*) that perform ever so slightly better than a random classifier. A natural question then arises (paraphrased from [Sch90]'s abstract): can one convert such mediocre learning rules into one that performs extremely well? Boosting algorithms offer a positive resolution to this question by providing convenient and computationally efficient reductions that aggregate such weak learners into a proficient learner with an arbitrarily high accuracy.

Realizable Boosting. Consider the celebrated Adaboost algorithm [FS97] which operates in the (noiseless) realizable binary classification setting. On any distribution consistent with a fixed labeling function (or *concept*), a weak learner here promises an accuracy strictly better than half, since guessing labels randomly gets any example correct with probability half. Given access to such a weak learner and $m_{\text{realizable boosting}} \approx (\log |\mathcal{H}|)/\varepsilon$ samples[1], Adaboost makes $\approx \log 1/\varepsilon$ calls to the weak learner to produce a classifier with (absolute) error at most ε on any distribution consistent with the same labeling function. Thus, not only is Adaboost computationally efficient provided, of course, such weak learners can be found, but also its sample complexity is no worse than that of ERM. *This underscores the fact that the realizable boosting methodology has the potential to offer computational relief, without compromising on sample efficiency.*

Agnostic Boosting. In practice, realizability is an onerous assumption; it is too limiting for the observed feature values alone to determine the label completely and deterministically, and that such a relation can be perfectly captured by an in-class hypothesis. Agnostic learning forgoes such assumptions. In their absence, bounds on absolute error are unachievable, e.g., when labels are uniformly random irrespective of features, no classifier can achieve accuracy better than half. Instead, in agnostic learning, the goal of the learner is to output a hypothesis with small *excess* error with respect to the best in-class classifier. If an in-class hypothesis is perfect on a given distribution, this relative error translates to an absolute error bound, thus generalizing the realizable case perfectly. Indeed, such model agnosticism has come to be a lasting hallmark of modern machine learning.

Early attempts at realizing the promise of boosting in the agnostic setting were met with limited success: while they did boost the weak learner's accuracy, the final hypothesis produced was not competitive with the best in-class hypothesis. We survey some of these in the related work section. A later result, and the work most related to ours, is due to [KK09]. A weak learner in this setting returns a classifier with a correlation against the true labels that is γ (say 0.1) times that of the best in-class

[1] In the introduction, for simplicity, we suppress polynomial dependencies in the weak learner's edge γ, and polylogarithmic terms. A recent sample complexity lower bound due to [GLR22] implies that this equivalence continues to hold even taking into account $\text{poly}(\gamma)$ dependencies as long as γ is not exponentially small.

	Episodic model	Rollouts w. ν-resets
[BHS22]	$1/\gamma^4\varepsilon^5$	$1/\gamma^4\varepsilon^6$
Theorem 7	$1/\gamma^3\varepsilon^4$	$1/\gamma^3\varepsilon^5$

Table 2: Sample complexity of reinforcement learning given γ-weak learner over the policy class, for two different modes of accessing the underlying MDP, in terms of ε and γ, suppressing other terms.

hypothesis. Random guesses of the labels produce a correlation of zero; hence, a weak classifier interpolates the performance of the best in-class hypothesis with that of a random one. Given access to such a weak learner, the boosting algorithm of [KK09] makes $\approx 1/\varepsilon^2$ calls to the weak learner and draws $m_{\text{agnostic boosting}} \approx (\log|\mathcal{H}|)/\varepsilon^4$ samples to produce a learning rule (not necessarily in the hypotheses class, hence *improper*) with ε-excess error. The dependency of the sample complexity in the target accuracy is thus quadratically worse than that of ERM. This gap persists untarnished for other known agnostic boosting algorithms too. *In this work, we seek to diminish this fundamental gap and construct a more sample-efficient agnostic boosting algorithm.*

Our main result is an efficient boosting algorithm that upon receiving $\approx (\log|\mathcal{H}|)/\varepsilon^3$ samples produces an *improper* learning rule with ε-excess error on the population loss. This is accomplished by the careful reuse of samples across rounds of boosting. We also extend these guarantees to infinite function classes, and give applications of our main result in reinforcement learning, and in agnostically learning halfspaces, in each case improving known results. We detail key contributions and technical challenges in achieving them next.

1.1 Contributions and technical innovations

Contribution 1: Sample-efficient Agnostic Booster. We provide a new potential-based agnostic boosting algorithm (Algorithm 1) that achieves ε-excess error when given a γ-weak learner operating with a base class $\mathcal{B}$. In Theorem 4, we prove that the sample complexity of this new algorithm scales as $(\log|\mathcal{B}|)/\gamma^3\varepsilon^3$, improving upon all known approaches to agnostic boosting. See Table 1.

A key innovation in our algorithm design and the source of our sample efficiency is the careful recursive reuse of samples between rounds of boosting, via a second-order estimator of the potential in Line 5.II. In contrast, [KK09] draws fresh samples for every round of boosting.

A second coupled innovation, this time in our analysis, is to circumvent a uniform convergence argument on *the boosted hypothesis class*, which would otherwise result in a sample complexity that scales as $\approx 1/\varepsilon^4$. Indeed, the algorithm in [BCHM20] reuses the entire training dataset across all rounds of boosting. This approach succeeds in boosting the accuracy on the *empirical* distribution; however, success on the *population* distribution now relies on a uniform convergence (or sample compression) argument on the boosted class, the complexity of which grows linearly with the number of rounds of boosting since boosting algorithms are inevitably improper (i.e., output a final hypothesis by aggregating weak learners, hence outside the base class). Instead, we use a martingale argument on *the smaller base hypothesis class* to show that the empirical distributions constructed by our data reuse scheme and fed to the weak learner track the performance of any base hypothesis on the population distribution. This is encapsulated in Lemma 6.

Finally, while we follow the potential-based framework laid in [KK09], we find it necessary to alter the branching logic dictating what gets added to the ensemble at every step. At each step, the algorithm makes progress via including the weak hypothesis or making a step "backward" via adding in a negation of the sign of the current ensemble. We note that there is a subtle error in [KK09] (see Appendix A), that although for their purposes is rectifiable without a change in the claimed sample complexity, leads to $1/\varepsilon^4$ sample complexity here in spite of the above modifications. At the leisure of $1/\varepsilon^4$ sample complexity, the fix is to test which of these alternatives fares better by drawing fresh samples every round. However, given a smaller budget, the error of the negation of the sign of the ensemble, which lies outside the base class, is not efficiently estimable. Instead, in Line 9 we give a different branching criteria that can be evaluated using the performance of the weak hypothesis on past data alone.

https://doi.org/10.52202/079017-3190

Contribution 2: Trading off Sample and Oracle Complexity. Although Theorem 4 offers an unconditional improvement on the sample complexity, it makes more calls to the weak learners than previous works. To rectify this, we give a second guarantee on the performance of Algorithm 1 in Theorem 9 (in Appendix B), with the oracle complexity matching that of known results. The resultant sample complexity improves known results for all practical (i.e., sub-constant) regimes of ε. This is made possible by a less well known variant of Freedman's inequality [Pin94, Pin20] that applies to random variables bounded with high probability.

Contribution 3: Extension to Infinite Classes. Although in our algorithm the samples fed to the weak leaner are independent conditioned on past sources of randomness, our relabeling and data reuse across rounds introduces complicated inter-dependencies between samples. For example, a sample drawn in the past can simultaneously be used as is in the present round (i.e., in Line 5.I), in addition to implicitly being used to modify the label of a different sample via the weak hypothesis it induced in the past (i.e., in Line 5.II). Thus, the textbook machinery of extending finite hypothesis results to infinite class applicable to IID samples via symmetrization and Rademacher complexity (see, e.g., [SSBD14, MRT18, BBL05]) is unavailable to us. Instead, we first derive L_1 covering number based bounds in Theorem 19 (in Appendix F). Through a result in empirical process theory [VDVW97], we translate these to a $\mathrm{VC}(\mathcal{B})/\gamma^3\varepsilon^3$ sample complexity bound, where $\mathrm{VC}(\mathcal{B})$ is the VC dimension of class $\mathcal{B}$, in Theorem 5.

Contribution 4: Applications in Reinforcement Learning and Agnostic Learning of Halfspaces. Building on earlier reductions from reinforcement learning to supervised learning [KL02], [BHS22] initiated the study of function-approximation compatible reinforcement learning, given access to a weak learner over the policy class. By applying our agnostic booster to this setting, we improve their sample complexity by $\mathrm{poly}(\varepsilon, \gamma)$ factors for binary-action MDPs, as detailed in Table 2. Also, following [KK09], we apply our agnostic boosting algorithm to the problem of learning halfspaces over the Boolean hypercube and exhibit improved boosting-based results in Theorem 8.

Contribution 5: Experiments. In preliminary experiments in Section 7, we demonstrate that the sample complexity improvements of our approach do indeed manifest in the form of improved empirical performance over previous agnostic boosting approaches on commonly used datasets.

2 Related work

The possibility of boosting was first posed in [KV94], and was resolved positively in a remarkable result due to [Sch90] for the realizable case. The Adaboost algorithm [FS97] paved the way for its practical applications (notably in [VJ01]). We refer the reader to [SF13] for a comprehensive text that surveys the many facets of boosting, including its connections to game theory and online learning. See also [HLR23, GLR22, AGHM21] for recent developments.

The fact that Adaboost and its natural variants are brittle in presence of label noise and lack of realizability [LS08] prompted the search for boosting algorithms in the realizable plus label noise [DIK+21, KS03] and agnostic learning models [BDLM01, MM02, KMV08, Kal07, CBC16]. In general, these boosting models are incomparable: although agnostic learning implies success in the random noise model, agnostic weak learning also constitutes a stronger assumption. Early agnostic boosting results could not boost the learner's accuracy to match that of the best in-class hypothesis; this limitation was tied to their notion of agnostic weak learning. Our work is most closely related to [KK09, Fel10]; we use the same notion of agnostic weak learning.

Boosting has also been extended to the online setting. [BKL15, CLL12, JGT17] study boosting in the mistake bound (realizable) model, while [BCHM20, RT22, HS21] focus on regret minimization. Our scheme of data reuse is inspired by variance reduction techniques [SLRB17, JZ13, FLLZ18, ABS23] in convex optimization, although there considerations of uniform convergence and generalization are absent, and our algorithm does not admit a natural gradient descent interpretation.

3 Problem setting

Let $\mathcal{D} \in \Delta(\mathcal{X} \times \{\pm 1\})$ be the joint population distribution over features, chosen from $\mathcal{X}$, and (signed) binary labels with respect to which a classifier's $h : \mathcal{X} \to \{\pm 1\}$ performance may be assessed. The

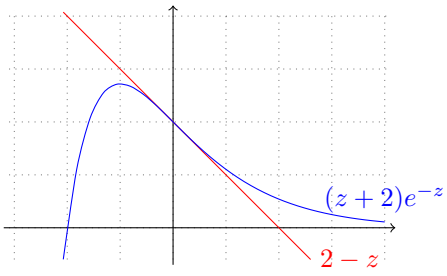

Figure 1: The two components of the piecewise potential function $\phi(z)$, with $(z+2)e^{-z}$ plotted in blue, and $2-z$ in red. Note that $\phi(z)$ is the point-wise maximum of the two.

performance criterion we consider is the 0-1 loss over the true labels and the classifier's predictions.

$$l_{\mathcal{D}}(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\mathbb{1}(h(x) \neq y)\right] = \Pr_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$$

Relatedly, one may measure the correlation between the classifier's predictions and the true labels.

$$\mathrm{corr}_{\mathcal{D}}(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[yh(x)\right]$$

Note that for signed binary labels, we have that $l_{\mathcal{D}}(h) = \frac{1}{2}(1 - \mathrm{corr}_{\mathcal{D}}(h))$. Therefore, these notions are equivalent in that a classifier that maximizes the correlation with true labels also minimizes the 0-1 loss, and vice versa, even in a relative error sense.

Definition 1 (Agnostic Weak Learner)**.** A learning algorithm is called a γ-agnostic weak learner with sample complexity $m : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{N} \cup \{+\infty\}$ with respect to a hypothesis class $\mathcal{H}$ and a base hypothesis class $\mathcal{B}$ if, for any $\varepsilon_0, \delta_0 > 0$, upon being given $m(\varepsilon_0, \delta_0)$ independently and identically distributed samples from any distribution $\mathcal{D}' \in \Delta(\mathcal{X} \times \{\pm 1\})$, it can output a base[2] hypothesis $\mathcal{W} \in \mathcal{B}$ such that with probability $1 - \delta_0$ we have

$$\mathrm{corr}_{\mathcal{D}'}(\mathcal{W}) \geq \gamma \max_{h\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}'}(h) - \varepsilon_0.$$

As remarked in [KK09], typically $m(\varepsilon, \delta) = O((\log|\mathcal{B}|/\delta)/\varepsilon^2)$, and we use this fact in compiling Table 1. However, following [KK09, BCHM20], we state our main result for fixed ε_0, δ_0 in Theorem 4, where a necessary and irreducible ε_0 term shows up in our final accuracy.

Although not explicitly mentioned in our weak learning definition, our algorithm falls within the *distribution-specific* boosting framework [KK09, Fel10]. In particular, like previous work on agnostic boosting, Algorithm 1 can be implemented by *relabeling examples*, instead of adaptively reweighing them. Thus, the overall marginal distribution of any $\mathcal{D}'$ fed to the learner on the feature space $\mathcal{X}$ is the same as that induced by the population distribution $\mathcal{D}$ on $\mathcal{X}$. Under such promise on inputs, distribution-specific weak learners may be easier to find.

4 The algorithm and main results

Notations. Given two functions $f, g : \mathcal{X} \to \mathbb{R}$ and generic scalars $\alpha, \beta \in \mathbb{R}$, we use $af + bg$ to denote a function such that $(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x)$ for all $x \in \mathcal{X}$. Given a function $f : \mathcal{X} \to \mathbb{R}$, we take $\mathrm{sign}(f)$ to be a function such that for all $x \in \mathcal{X}$, $\mathrm{sign}(f)(x) = \mathbb{1}(f(x) \geq 0) - \mathbb{1}(f(x) < 0)$. Define a filtration sequence $\{\mathcal{F}_t : t \in \mathbb{N}_{\geq 0}\}$, where $\mathcal{F}_t$ capture all source of randomness the algorithm is subject to in the first t iterations. For brevity, we define $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$. For any feature-label dataset $\widehat{D}$, we use $\mathbb{E}_{\widehat{D}}$ and $\mathrm{corr}_{\widehat{D}}$ to denote the empirical average and empirical correlation over $\widehat{D}$.

Potential function. Define the potential function $\phi : \mathbb{R} \to \mathbb{R}$ as

$$\phi(z) = \begin{cases} 2 - z & \text{if } z \leq 0, \\ (z+2)e^{-z} & \text{if } z > 0. \end{cases} \tag{1}$$

[2]Typically, the base class $\mathcal{B}$ is (often substantially) smaller than $\mathcal{H}$. For example, decision stumps are a common example of the base class.

https://doi.org/10.52202/079017-3190

Algorithm 1 Agnostic Boosting via Sample Reuse

Inputs: Sampling oracle for $\mathcal{D}$ supported on $\mathcal{X} \times \{\pm 1\}$, γ-agnostic weak learning oracle $\mathcal{W}$, step-size η, mixing parameter σ, number of iterations T, per-iteration sample size S, resampling parameter m, branching tolerance τ, post-selection sample size S_0, potential $\phi : \mathbb{R} \to \mathbb{R}$.
Initialize a zero hypothesis $H_1 = \mathbf{0}$.
for $t = 1$ to T **do**
  Sample S IID examples from the distribution $\mathcal{D}$ to create a dataset $\widehat{D}_t$.
  Construct a sampling distribution $\mathcal{D}_t$ that samples (x, y) uniformly from $\widehat{D}_t$ if $t = 1$, and for $t > 1$ produces IID samples $(x, \hat{y})$ as follows:
    I With probability $1 - \sigma$, return a sample $(x, \hat{y})$ from $\mathcal{D}_{t-1}$.
    II With remaining probability σ, draw $\eta' \sim \mathrm{Unif}[0, \eta]$, pick (x, y) uniformly from $\widehat{D}_t$, construct a pseudo label $\hat{y} = \begin{cases} +1 & \text{with probability } p_t(x, y, \eta'), \\ -1 & \text{with probability } 1 - p_t(x, y, \eta'), \end{cases}$ and return $(x, \hat{y})$, where

$$p_t(x, y, \eta') = \frac{1}{2} - \frac{\sigma\phi'(yH_{t-1}(x))y + \eta\phi''(y(H_{t-1}(x) + \eta' h_{t-1}(x)))h_{t-1}(x)}{2(\eta + \sigma)}.$$

  Sample m samples from $\mathcal{D}_t$ to create another dataset $\widehat{D}'_t$.
  Call the weak learning oracle on $\widehat{D}'_t$ to get $W_t = \mathcal{W}(\widehat{D}'_t)$.
  Measure the empirical correlation of W_t on $\widehat{D}'_t$ as $\mathrm{corr}_{\widehat{D}'_t}(W_t) = \sum_{(x,\hat{y})\in\widehat{D}'_t} \hat{y} W_t(x)$.
  Set $h_t = W_t/\gamma$ if $\mathrm{corr}_{\widehat{D}'_t}(W_t) > \tau$ else $h_t = -\operatorname{sign}(H_t)$.
  Update $H_{t+1} = H_t + \eta h_t$.
end for
Sample S_0 IID examples from the distribution $\mathcal{D}$ to create a dataset $\widehat{D}_0$.
Output the hypothesis $\overline{h} = \arg\max_{h\in\{\operatorname{sign}(H_t):t\in[T]\}} \sum_{(x,y)\in\widehat{D}_0} yh(x)$.

We can use this to assign a population potential to any real-valued hypothesis $H : \mathcal{X} \to \mathbb{R}$ as

$$\Phi_{\mathcal{D}}(H) = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi(yH(x)\right].$$

To maximize correlation between true labels and the hypothesis H's outputs, one wants $H(x) > 0$ whenever $y = +1$ for most samples drawn from the underlying distribution. Since ϕ is a monotonically decreasing function, equivalently, higher classifier accuracy typically corresponds to lower values of population potential. However, there are limits to the utility of this argument: a low value of the potential alone does not translate to successful learning. In agnostic learning, one is concerned not with the error of the learned classifier *per se*, but with its excess error over the best in-class hypothesis. We will provide a precise relation between the potential and the excess error in Lemma 3.

We will use the following properties of ϕ (proved in Appendix C). Going forward, these will be the sole characteristics of ϕ we will appeal to. The potential we use is similar to the one used in [KK09, Dom00], but has been modified to remove a jump discontinuity in the second derivative at $z = 0$ as our approach requires a twice continuously differentiable potential.

Lemma 2. *We make the following elementary observations about ϕ:*

I. ϕ is convex and in $\mathcal{C}^2$, i.e., it is two-times continuously differentiable everywhere.
II. ϕ is non-negative on $\mathbb{R}$ and $\phi(0) = 2$.
III. For all $z \in \mathbb{R}$, $\phi'(z) \in [-1, 0]$. Further, for any $z < 0$, $\phi'(z) = -1$.
IV. ϕ is 1-smooth, i.e., $\forall z \in \mathbb{R}$, $\phi''(z) \leq 1$.

For any real-valued hypotheses H, h and g on $\mathcal{X}$, to ease analysis, we introduce

$$\Phi'_{\mathcal{D}}(H, h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\phi'(yH(x))yh(x)],$$
$$\Phi''_{\mathcal{D}}(H, h, g) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\phi''(yH(x))h(x)g(x)].$$

Equivalently, $\Phi'_{\mathcal{D}}(H, h)$ can be characterized as the $\mathcal{D}$-induced semi inner product between the functional derivative $\partial\Phi_{\mathcal{D}}(H)/\partial H$ and h. But for ease of presentation, we forgo this formal interpretation in favor of the literal one stated above.

A key property of the above potential is stated in the next lemma (proved in Appendix C). It gives us a strategy to control the relative error of the learned hypothesis, the quantity on the right, by individually minimizing both terms on the left, as we discuss next.

Lemma 3. *For any distribution $\mathcal{D} \in \Delta(\mathcal{X} \times \{\pm 1\})$, real-valued hypothesis $H : \mathcal{X} \to \mathbb{R}$, and binary hypothesis $h^* : \mathcal{X} \to \{\pm 1\}$, we have*

$$\Phi'_{\mathcal{D}}(H, \operatorname{sign}(H)) - \Phi'_{\mathcal{D}}(H, h^*) \geq \operatorname{corr}_{\mathcal{D}}(h^*) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign}(H)).$$

Description of the algorithm. Each round of Algorithm 1 adds some multiple of either the weak hypothesis W_t or $-\operatorname{sign}(H_t)$ to the current ensemble H_t; this choice is dictated by the empirical correlation of the weak hypothesis on the dataset it was fed. The construction of the relabeled distribution $\mathcal{D}_t$ via our data reuse scheme ensures that if any hypothesis in the base class $\mathcal{B}$ produces sufficient correlation on it, its addition to the ensemble must decrease the potential Φ associated with the ensemble. Concretely, as we prove in Lemma 6, for all $h \in \mathcal{B}$, $\operatorname{corr}_{\mathcal{D}_t}(h)$ closely tracks $-\Phi'(H_t, h)$. The key to our improved sample complexity is the fact that this invariant can be maintained while sampling only $S \approx 1/\gamma\varepsilon$ fresh samples each round, by repurposing samples from earlier rounds to construct the dataset fed to the weak learner. The mixing parameter σ controls the proportion of these two sources of samples used to construct $\mathcal{D}_t$.

However, this explanation is opaque when it comes to motivating the need for $-\operatorname{sign}(H_t)$. Let's rectify that: let $h^* \in \arg\min_{h \in \mathcal{H}} \operatorname{corr}_{\mathcal{D}}(h)$ be the best in-class hypothesis. If $-\Phi'_{\mathcal{D}}(H_t, h^*)$ is sufficiently large, so is $\operatorname{corr}_{\mathcal{D}_t}(h^*)$ by Lemma 6, which assures us a non-negligible weak learning edge. As such $-\Phi'_{\mathcal{D}}(H_t, W_t)$ is large and the potential drops. If the weak hypothesis fails to make sufficient progress, the algorithms adds $-\operatorname{sign}(H_t)$ to the ensemble, which, again by Lemma 6 corresponds to decreasing the potential value by some function of $-\Phi'_{\mathcal{D}}(H_t, -\operatorname{sign}(H_t)) = \Phi'_{\mathcal{D}}(H_t, \operatorname{sign}(H_t))$, using linearity of $\Phi'_{\mathcal{D}}$ in its second argument. Thus, when run for sufficiently many iterations, because the potential is bounded above at initialization both the terms on the left side of Lemma 3 must become small. This then implies a bounded correlation gap between the best in-class hypothesis, and the majority vote of the ensembles considered in some iteration; the last line picks the best of these.

4.1 Main result for finite hypotheses classes

The key feature of our algorithm is that it is designed to reuse samples across successive rounds of boosting. The soundness of this scheme, which Lemma 6 substantiates, is based on the observation that in each round H_t changes by a small amount, thereby inducing an incremental change in the distribution fed to the weak learner. Although our algorithm needs a total number of iterations comparable to [KK09], this data reuse lowers the number of fresh samples needed every iteration to just $1/\gamma\varepsilon$, instead of $1/\gamma^2\varepsilon^2$, resulting in improved sample complexity, as we show next.

Theorem 4 (Main result for finite hypotheses class)**.** *Choose any $\varepsilon, \delta > 0$. There exists a choice of $\eta, \sigma, T, \tau, S_0, S, m$ satisfying[3] $T = \mathcal{O}((\log|\mathcal{B}|)/\gamma^2\varepsilon^2), \eta = \mathcal{O}(\gamma^2\varepsilon/\log|\mathcal{B}|), \sigma = \eta/\gamma, \tau = \mathcal{O}(\gamma\varepsilon), S = \mathcal{O}(1/\gamma\varepsilon), S_0 = \mathcal{O}(1/\varepsilon^2), m = m(\varepsilon_0, \delta_0) + \mathcal{O}(1/\gamma^2\varepsilon^2)$ such that for any γ-agnostic weak learning oracle (as defined in Definition 1) with fixed tolerance ε_0 and failure probability δ_0, Algorithm 1 when run with the potential defined in (1) produces a hypothesis $\overline{h}$ such that with probability $1 - 10\delta_0 T - 10\delta T$,*

$$\operatorname{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h \in \mathcal{H}} \operatorname{corr}_{\mathcal{D}}(h) - \frac{2\varepsilon_0}{\gamma} - \varepsilon,$$

while making $T = \mathcal{O}((\log|\mathcal{B}|)/\gamma^2\varepsilon^2)$ calls to the weak learning oracle, and sampling $TS + S_0 = \mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^3)$ labeled examples from $\mathcal{D}$.

In Appendix B, we provide a different result (Theorem 9), also using Algorithm 1, where the learner makes $\mathcal{O}(1/\gamma^2\varepsilon^2)$ call to weak learner, exactly matching the oracle complexity of existing results, while drawing $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^3 + (\log|\mathcal{B}|)^3/\gamma^2\varepsilon^2)$ samples.

4.2 Extensions to infinite classes

As mentioned in Section 1.1, the reuse of samples prohibits us from appealing to symmetrization and Rademacher complexity based arguments. Instead, our generalization to infinite classes is based

[3] Here, the $\mathcal{O}$ notation suppresses polylogarithmic factors.

https://doi.org/10.52202/079017-3190

on L_1 covering numbers. Using empirical process theory [VDVW97], we upper bound L_1 covering number by a suitable function of VC dimension, yielding the following result (proved in Appendix F).

Theorem 5 (Main result for VC Dimension)**.** *There exists a setting of parameters such that for any for any γ-agnostic weak learning oracle with fixed tolerance ε_0 and failure probability δ_0, Algorithm 1 produces a hypothesis $\overline{h}$ such that with probability $1 - 10\delta_0 T - 10\delta T$,*

$$\operatorname{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h \in \mathcal{H}} \operatorname{corr}_{\mathcal{D}}(h) - \frac{2\varepsilon_0}{\gamma} - \varepsilon,$$

while making $T = \mathcal{O}(VC(\mathcal{B})/\gamma^2\varepsilon^2)$ calls to the weak learning oracle, and sampling $TS + S_0 = \mathcal{O}(VC(\mathcal{B})/\gamma^3\varepsilon^3)$ labeled examples from $\mathcal{D}$.

5 Sketch of the analysis

In this section, we provide a brief sketch of the analysis outlined in the proof of Theorem 4. Our intent is to convey the plausibility of a $1/\varepsilon^3$ sample complexity result, up to the exclusion of other factors. Hence, $\approx$ and $\lesssim$ inequalities below only hold up to constants and polynomial factors in other paramters, e.g., in $\gamma, \log|\mathcal{B}|$. Formal proofs are reserved for Appendix D.

Bounding the correlation gap (Theorem 4). A central tool in bounding the correlation gap is Lemma 3. We want to ensure for some t, since our algorithm at the end picks the best one, that

$$\underbrace{-\Phi'_{\mathcal{D}}(H_t, -\operatorname{sign}(H_t))}_{\text{want } \lesssim \varepsilon} + \underbrace{(-\Phi'_{\mathcal{D}}(H_t, h^*))}_{\text{want } \lesssim \varepsilon} \geq \operatorname{corr}_{\mathcal{D}}(h^*) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign}(H_t)).$$

On the other hand, using 1-smoothness of ϕ, we can upper bound $\Phi_{\mathcal{D}}$ on successive iterates as $\Phi_{\mathcal{D}}(H_{t+1}) \leq \Phi_{\mathcal{D}}(H_t) + \eta\Phi'_{\mathcal{D}}(H_t, h_t) + \eta^2/2\gamma^2$. Rearranging this to telescope the sum produces

$$-\frac{1}{T}\sum_{t=1}^{T} \Phi'_{\mathcal{D}}(H_t, h_t) \leq \frac{\sum_{t=1}^{T}(\Phi_{\mathcal{D}}(H_t) - \Phi_{\mathcal{D}}(H_{t+1}))}{\eta T} + \frac{\eta}{2\gamma^2} \leq \frac{2}{\eta T} + \frac{\eta}{2\gamma^2}$$

Hence, by setting a $\eta \approx 1/\sqrt{T}$, we know that there exists a t where $-\Phi'_{\mathcal{D}}(H_t, h_t) \lesssim 1/\sqrt{T}$.

In Lemma 6, we establish that the core guarantee our data resue scheme provides: that for all h in the base class $\mathcal{B}$ and for h^*, the correlation on the resampled distribution $\mathcal{D}_t$ constructed by the algorithm tracks the previously stated quantity of interest $-\Phi'_{\mathcal{D}}(H_t, \cdot)$.

Lemma 6. *There exists a $C > 0$ such that with probability $1 - \delta$, for all $t \in [T]$ and $h \in \mathcal{B} \cup \{h^*\}$, we have*

$$\left|\Phi'_{\mathcal{D}}(H_t, h) + \left(1 + \frac{\eta}{\sigma}\right)\operatorname{corr}_{\mathcal{D}_t}(h)\right| \leq \underbrace{C\left(\sigma + \frac{\eta}{\gamma}\right)\left(\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{|\mathcal{B}|T}{\delta}} + \log\frac{|\mathcal{B}|T}{\delta}\right)}_{:=\varepsilon_{Gen}}.$$

Using the definition of weak learner, we know that $\operatorname{corr}_{\mathcal{D}_t}(h^*) \lesssim \operatorname{corr}_{\mathcal{D}_t}(W_t)/\gamma$. Now, using Lemma 6 twice and the linearity of $\Phi'_{\mathcal{D}}(H, \cdot)$, we get

$$-\Phi'_{\mathcal{D}}(H_t, h^*) \lesssim \operatorname{corr}_{\mathcal{D}_t}(h^*) + \varepsilon_{\text{Gen}} \lesssim \operatorname{corr}_{\mathcal{D}_t}(W_t)/\gamma + \varepsilon_{\text{Gen}} \lesssim -\Phi'_{\mathcal{D}}(H_t, W_t/\gamma) + 2\varepsilon_{\text{Gen}}.$$

Now, if at each step we could choose $h_t \in \{-\operatorname{sign}(H_t), W_t/\gamma\}$, which ever maximized $-\Phi'_{\mathcal{D}}(H_t, \cdot)$, we would have for some t that

$$\max\{-\Phi'_{\mathcal{D}}(H_t, -\operatorname{sign}(H_t)), -\Phi'_{\mathcal{D}}(H_t, h^*) - 2\varepsilon_{\text{Gen}}\} \leq -\Phi'_{\mathcal{D}}(H_t, h_t) \lesssim 1/\sqrt{T}.$$

Alas, $-\operatorname{sign}(H_t)$ is not in $\mathcal{B}$, thus $\operatorname{corr}_{\mathcal{D}_t}(-\operatorname{sign}(H_t))$ can be really far from $-\Phi'_{\mathcal{D}}(H_t, -\operatorname{sign}(H_t))$, i.e., Lemma 6 does not apply to $-\operatorname{sign}(H_t)$. To circumvent this, instead of choosing the maximizer, in the algorithm and in the actual proof, we use a relaxed criteria for choosing between W_t/γ and $-\operatorname{sign}(H_t)$ that depends on the correlation of W_t on $\mathcal{D}_t$ alone, and hence can be efficiently evaluated. The spirit of this modification is to adopt $-\operatorname{sign}(H_t)$ only if W_t by itself fails to make enough progress, the threshold for which can be stated in the terms of target accuracy.

Generalization over $\mathcal{B}$ via sample reuse (Lemma 6). Here, we sketch a proof of Lemma 6 that ties the previous proof sketch together, and forms the basis of our sample reuse scheme. Our starting point is the following claim (Claim 10) that uses the fact that ϕ is second-order differentiably continuous to arrive at the fact that for any t and $h : \mathcal{X} \to \mathbb{R}$, we have

$$\Phi'_{\mathcal{D}}(H_t, h) \approx \Phi'_{\mathcal{D}}(H_{t-1}, h) + \eta \Phi''_{\mathcal{D}}(H_{t-1}, h, h_{t-1}).$$

Simultaneously, using $\mathbb{E}_{t-1}$ to condition on the randomness in the first $t-1$ rounds, by construction, our data reuse and relabeling scheme gives us

$$\mathbb{E}_{t-1}[\mathrm{corr}_{\mathcal{D}_t}(h)] \approx (1-\sigma)\mathrm{corr}_{\mathcal{D}_{t-1}}(h) - \frac{\sigma^2}{\sigma+\eta}\Phi'_{\mathcal{D}}(H_{t-1}, h) - \frac{\eta\sigma}{\sigma+\eta}\Phi''_{\mathcal{D}}(H_{t-1}, h, h_{t-1}).$$

Thus, suitably scaling and adding the two together, we get the identity that

$$\begin{aligned}\mathbb{E}_{t-1}[\Delta_t] &= \Phi'_{\mathcal{D}}(H_t, h) + \left(1 + \frac{\eta}{\sigma}\right)\mathbb{E}_{t-1}\left[\mathrm{corr}_{\mathcal{D}_t}(h)\right] \\ &= (1-\sigma)\Phi'_{\mathcal{D}}(H_{t-1}, h) + (1-\sigma)\left(1+\frac{\eta}{\sigma}\right)\mathrm{corr}_{\mathcal{D}_{t-1}}(h) = (1-\sigma)\Delta_{t-1},\end{aligned}$$

where $\Delta_t = \Phi'_{\mathcal{D}}(H_t, h) + \left(1+\frac{\eta}{\sigma}\right)\mathrm{corr}_{\mathcal{D}_t}(h)$. Thus, Δ_t forms a martingale-like sequence; the sign of Δ_t is indeterminate. To establish concentration, we apply Freedman's inequality, noting that the conditional variance of the associated martingale difference sequence scales as $1/\sqrt{S}$. A union bound over $\mathcal{B} \cup \{h^*\}$ yields the claim. Relatedly, to reach a better oracle complexity in Theorem 9, we show a uniform high-probability on the martingale difference sequence, and then apply a variant of Freedman's inequality [Pin94, Pin20] that can adapt to martingale difference sequences that are bounded with high probability instead of almost surely.

6 Applications

In this section, we detail the implications of our results for previously studied learning problems.

6.1 Boosting for reinforcement learning

[BHS22] initiated the approach of boosting weak learners to construct a near-optimal policy for reinforcement learning. Plugging Algorithm 1 into their meta-algorithm yields the following result for binary-action MDPs, improving upon the sample complexity in [BHS22]. Here, V^π is the expected discounted reward of a policy π, V^* is its maximum. β is the discount factor of the underlying MDP, and C_∞, D_∞ and $\mathcal{E}, \mathcal{E}_\nu$ are distribution mismatch and policy completeness terms (related to the inherent Bellman error). In the *episodic model*, the learner interacts with the MDP in episodes. In the *ν-reset model*, the learner can seed the initial state with a fixed well dispersed distribution ν as a means to exploration. See Appendix G for a complete statement of results and details of the setting.

Theorem 7 (Informal; stated formally in Theorem 22)**.** *Let $\mathcal{W}$ be a γ-weak learner for the policy class Π operating with a base class $\mathcal{B}$, with sample complexity $m(\varepsilon_0, \delta_0) = (\log|\mathcal{B}|/\delta_0)/\varepsilon_0^2$. Fix tolerance ε and failure probability δ. In the* episodic *access model, there is an algorithm using that uses the weak learner $\mathcal{W}$ to produce a policy $\overline{\pi}$ such that with probability $1-\delta$, we have $V^* - V^{\overline{\pi}} \le (C_\infty \mathcal{E})/(1-\beta) + \varepsilon$, while sampling $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^4)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$. In the ν-*reset *access model, there is a setting of parameters such that Algorithm 2 when given access to $\mathcal{W}$ produces a policy $\overline{\pi}$ such that with probability $1-\delta$, we have $V^* - V^{\overline{\pi}} \le (D_\infty \mathcal{E}_\nu)/(1-\beta)^2 + \varepsilon$, while sampling $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^5)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$.*

6.2 Agnostically learning halfspaces

We apply our algorithm in a black-box manner to agnostically learn halfspaces over the n-dimensional boolean hypercube when the data distribution has uniform marginals on features. The aim is this section is not to obtain the best known bounds, but rather to provide an example illustrating that agnostic boosting is both a viable and flexible approach to construct agnostic learners, and where our improvements carry over. Following [KK09], we use ERM over the parities of degree at most d, for $d \approx 1/\varepsilon^4$, as our weak learners; the the weak learner's edge here is $\gamma = n^{-d}$. An application of our boosting algorithm (proved in Appendix I) to this problem improves the sample complexity of $\mathcal{O}(\varepsilon^{-8} n^{80\varepsilon^{-4}})$ indicated in [KK09].

https://doi.org/10.52202/079017-3190

Theorem 8. *Let $\mathcal{D}$ be any distribution over $\{\pm 1\}^n \times \{\pm 1\}$ with uniform distribution over features. By $\mathcal{H} = \{\mathrm{sign}(w^\top x - \theta) : (w, \theta) \in \mathbb{R}^{n+1}\}$, denote the class of halfspaces. There exists some d such that running Algorithm 1 with ERM over parities of degree at most d produces a classifier $\bar{h}$ such that $l_{\mathcal{D}}(\bar{h}) \leq \min_{h \in \mathcal{H}} l_{\mathcal{D}}(h) + \varepsilon$, while using $\mathcal{O}(\varepsilon^{-7} n^{60\varepsilon^{-4}})$ samples in $n^{poly(1/\varepsilon)}$ time.*

Note that ERM over the class of halfspaces directly, although considerably more sample efficient, takes $(1/\varepsilon)^{\mathrm{poly}(n)}$ time, i.e., it is exponentially slower for moderate values of ε. There are known statistical query lower bounds [DKZ20] requiring $n^{\mathrm{poly}(\varepsilon^{-1})}$ queries for agnostic learning of halfspaces with Gaussian marginals, suggesting that a broad class of algorithms, regardless of the underlying parametrization, might not fare any better. For completeness, we note that a better sample complexity is attainable by direct L_1-approximations of halfspaces via low-degree polynomials [DGJ+10], instead of the approach taken in [KK09, KOS04] and mirrored here which first constructs an L_2-approximation, but such structural improvements apply equally to presented and compared results.

7 Experiments

In Table 3, we report the results of preliminary experiments with Algorithm 1 against the agnostic boosting algorithms in [KK09] and [BCHM20] as baselines on UCI classification datasets [SWHB89, HRFS99, SED+88], using decision stumps [PVG+11] as weak learners. We also introduce classification noise of 5%, 10% and 20% during training to measure the robustness of the algorithms to label noise. Accuracy is estimated using 30-fold cross validation with a grid search over the mixing weight σ and the number of boosters T. The algorithm in [KK09] does not reuse samples between rounds, while [BCHM20] uses the same set of samples across all rounds. In contrast, Algorithm 1 blends fresh and old samples every round, with σ controlling the proportion of each. See Appendix J for additional details. We note that the Ionsphere dataset includes 351 samples, while Diabetes contains 768, and Spambase contains 4601. The benefits of sample reuse are less stark in a data-rich regime. This could explain some of the under-performance on Spambase, disregarding which the proposed algorithm substantially outperforms the alternatives.

Dataset	No Added Noise			5% Noise		
	[KK09]	[BCHM20]	**Ours**	[KK09]	[BCHM20]	**Ours**
Ionosphere	0.92 ± 0.02	0.89 ± 0.03	$\mathbf{0.97 \pm 0.02}$	0.90 ± 0.03	0.88 ± 0.03	$\mathbf{0.97 \pm 0.03}$
Diabetes	0.83 ± 0.03	0.78 ± 0.02	$\mathbf{0.87 \pm 0.03}$	0.83 ± 0.03	0.77 ± 0.02	$\mathbf{0.88 \pm 0.03}$
Spambase	0.69 ± 0.02	$\mathbf{0.79 \pm 0.01}$	0.78 ± 0.02	$\mathbf{0.81 \pm 0.02}$	0.79 ± 0.01	0.78 ± 0.02
German	0.77 ± 0.02	0.75 ± 0.02	$\mathbf{0.83 \pm 0.02}$	0.78 ± 0.02	0.75 ± 0.02	$\mathbf{0.85 \pm 0.02}$
Sonar	0.66 ± 0.07	$\mathbf{0.91 \pm 0.03}$	0.88 ± 0.07	0.84 ± 0.05	0.88 ± 0.03	$\mathbf{0.94 \pm 0.05}$
Waveform	0.88 ± 0.01	0.78 ± 0.01	$\mathbf{0.91 \pm 0.01}$	0.88 ± 0.01	0.77 ± 0.01	$\mathbf{0.90 \pm 0.01}$
Dataset	10% Noise			20% Noise		
	[KK09]	[BCHM20]	**Ours**	[KK09]	[BCHM20]	**Ours**
Ionosphere	0.93 ± 0.02	0.89 ± 0.02	$\mathbf{0.97 \pm 0.02}$	0.92 ± 0.03	0.90 ± 0.03	$\mathbf{0.96 \pm 0.03}$
Diabetes	0.83 ± 0.03	0.78 ± 0.02	$\mathbf{0.88 \pm 0.03}$	0.82 ± 0.02	0.78 ± 0.02	$\mathbf{0.88 \pm 0.02}$
Spambase	$\mathbf{0.83 \pm 0.02}$	0.80 ± 0.01	0.79 ± 0.02	$\mathbf{0.84 \pm 0.01}$	0.79 ± 0.01	0.79 ± 0.01
German	0.78 ± 0.02	0.75 ± 0.02	$\mathbf{0.84 \pm 0.02}$	0.78 ± 0.02	0.74 ± 0.02	$\mathbf{0.84 \pm 0.02}$
Sonar	0.85 ± 0.04	$\mathbf{0.91 \pm 0.03}$	0.88 ± 0.04	0.88 ± 0.04	0.88 ± 0.04	$\mathbf{0.93 \pm 0.04}$
Waveform	0.88 ± 0.01	0.77 ± 0.01	$\mathbf{0.91 \pm 0.01}$	0.88 ± 0.01	0.77 ± 0.01	$\mathbf{0.90 \pm 0.01}$

Table 3: Cross-validated accuracies of Algorithm 1 compared to the agnostic boosting algorithms from [KK09] and [BCHM20] on 6 datasets. The first column reports accuracy on the original datasets, and the next three report performance with 5%, 10% and 20% label noise added during training. The proposed algorithm simultaneously outperforms both the alternatives on 18 out of 24 instances.

8 Conclusion

We give an agnostic boosting algorithm with a substantially lower sample requirement than ones known, enabled by efficient recency-aware data reuse between boosting iterations. Improving our oracle complexity or proving its optimality, and closing the sample complexity gap to ERM are interesting directions for future work.

References

[ABS23] Naman Agarwal, Brian Bullins, and Karan Singh. Variance-reduced conservative policy iteration. In Shipra Agrawal and Francesco Orabona, editors, *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201 of *Proceedings of Machine Learning Research*, pages 3–33. PMLR, 20 Feb–23 Feb 2023.

[AGHM21] Noga Alon, Alon Gonen, Elad Hazan, and Shay Moran. Boosting simple learners. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 481–489, 2021.

[BBL05] Stéphane Boucheron, Olivier Bousquet, and Gábor Lugosi. Theory of classification: A survey of some recent advances. *ESAIM: probability and statistics*, 9:323–375, 2005.

[BCHM20] Nataly Brukhim, Xinyi Chen, Elad Hazan, and Shay Moran. Online agnostic boosting via regret minimization. *Advances in Neural Information Processing Systems*, 33:644–654, 2020.

[BDLM01] Shai Ben-David, Philip M Long, and Yishay Mansour. Agnostic boosting. In *Computational Learning Theory: 14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001 Amsterdam, The Netherlands, July 16–19, 2001 Proceedings 14*, pages 507–516. Springer, 2001.

[BHS22] Nataly Brukhim, Elad Hazan, and Karan Singh. A boosting approach to reinforcement learning. *Advances in Neural Information Processing Systems*, 35:33806–33817, 2022.

[BKL15] Alina Beygelzimer, Satyen Kale, and Haipeng Luo. Optimal and adaptive algorithms for online boosting. In *International Conference on Machine Learning*, pages 2323–2331. PMLR, 2015.

[CBC16] Shang-Tse Chen, Maria-Florina Balcan, and Duen Horng Chau. Communication efficient distributed agnostic boosting. In *Artificial Intelligence and Statistics*, pages 1299–1307. PMLR, 2016.

[CLL12] Shang-Tse Chen, Hsuan-Tien Lin, and Chi-Jen Lu. An online boosting algorithm with theoretical justifications. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pages 1873–1880, 2012.

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010.

[DIK+21] Ilias Diakonikolas, Russell Impagliazzo, Daniel M Kane, Rex Lei, Jessica Sorrell, and Christos Tzamos. Boosting in the presence of massart noise. In *Conference on Learning Theory*, pages 1585–1644. PMLR, 2021.

[DKZ20] Ilias Diakonikolas, Daniel Kane, and Nikos Zarifis. Near-optimal sq lower bounds for agnostically learning halfspaces and relus under gaussian marginals. *Advances in Neural Information Processing Systems*, 33:13586–13596, 2020.

[Dom00] C Domingo. Madaboost: a modification of adaboost. In *Proc. of the 13th Conference on Computational Learning Theory, COLT'00*, 2000.

[Dud74] Richard M Dudley. Metric entropy of some classes of sets with differentiable boundaries. *Journal of Approximation Theory*, 10(3):227–236, 1974.

[Fel10] Vitaly Feldman. Distribution-specific agnostic boosting. *Innovations in Theoretical Computer Science (ITCS)*, pages 241–250, 2010.

[FLLZ18] Cong Fang, Chris Junchi Li, Zhouchen Lin, and Tong Zhang. Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator. *Advances in neural information processing systems*, 31, 2018.

 https://doi.org/10.52202/079017-3190

[Fre75] David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

[FS97] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[GLR22] Kasper Green Larsen and Martin Ritzert. Optimal weak to strong learning. *Advances in Neural Information Processing Systems*, 35:32830–32841, 2022.

[Hau95] David Haussler. Sphere packing numbers for subsets of the boolean n-cube with bounded vapnik-chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69(2):217–232, 1995.

[HLR23] Mikael Møller Høgsgaard, Kasper Green Larsen, and Martin Ritzert. AdaBoost is not an optimal weak to strong learner. In *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 13118–13140. PMLR, 2023.

[HRFS99] Mark Hopkins, Erik Reeber, George Forman, and Jaap Suermondt. Spambase. UCI Machine Learning Repository, 1999. DOI: https://doi.org/10.24432/C53G6X.

[HS21] Elad Hazan and Karan Singh. Boosting for online convex optimization. In *International Conference on Machine Learning*, pages 4140–4149. PMLR, 2021.

[JGT17] Young Hun Jung, Jack Goetz, and Ambuj Tewari. Online multiclass boosting. *Advances in neural information processing systems*, 30, 2017.

[JZ13] Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in neural information processing systems*, 26, 2013.

[Kal07] Satyen Kale. Boosting and hard-core set constructions: a simplified approach. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14. Citeseer, 2007.

[KK09] Varun Kanade and Adam Kalai. Potential-based agnostic boosting. *Advances in neural information processing systems*, 22, 2009.

[KL02] Sham Kakade and John Langford. Approximately optimal approximate reinforcement learning. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pages 267–274, 2002.

[KMV08] Adam Tauman Kalai, Yishay Mansour, and Elad Verbin. On agnostic boosting and parity learning. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 629–638, 2008.

[KOS04] Adam R Klivans, Ryan O'Donnell, and Rocco A Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer and System Sciences*, 68(4):808–840, 2004.

[KS03] Adam Kalai and Rocco A Servedio. Boosting in the presence of noise. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 195–205, 2003.

[KV94] Michael Kearns and Leslie Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, 41(1):67–95, 1994.

[LS08] Philip M Long and Rocco A Servedio. Random classification noise defeats all convex potential boosters. In *Proceedings of the 25th international conference on Machine learning*, pages 608–615, 2008.

[MM02] Yishay Mansour and David McAllester. Boosting using branching programs. *Journal of Computer and System Sciences*, 64(1):103–112, 2002.

[MRT18] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[Pin94] Iosif Pinelis. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pages 1679–1706, 1994.

[Pin20] Iosif Pinelis. Extension of bernstein's inequality when the random variable is bounded with large probability. MathOverflow, 2020. URL:https://mathoverflow.net/q/371436 (version: 2020-09-11).

[Put14] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[RT22] Vinod Raman and Ambuj Tewari. Online agnostic multiclass boosting. *Advances in Neural Information Processing Systems*, 35:25908–25920, 2022.

[Sch90] Robert E Schapire. The strength of weak learnability. *Machine learning*, 5:197–227, 1990.

[SED+88] Jack W Smith, James E Everhart, WC Dickson, William C Knowler, and Robert Scott Johannes. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the annual symposium on computer application in medical care*, page 261. American Medical Informatics Association, 1988.

[SF13] Robert E Schapire and Yoav Freund. Boosting: Foundations and algorithms. *Kybernetes*, 42(1):164–166, 2013.

[SLRB17] Mark Schmidt, Nicolas Le Roux, and Francis Bach. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162:83–112, 2017.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[SWHB89] V. Sigillito, S. Wing, L. Hutton, and K. Baker. Ionosphere. UCI Machine Learning Repository, 1989. DOI: https://doi.org/10.24432/C5W01B.

[VDVW97] Aad W Van Der Vaart and Jon A Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer New York, 1997.

[VJ01] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001.

Appendix

Limitations. The primary contribution of this work is theoretical. Extensively demonstrating the empirical efficacy of the proposed approach and fully characterizing when it fares best is left to future work. Further, in comparison to realizable boosting (with or without label noise), the existence of an agnostic weak leaner here is a stronger assumption. Nevertheless, we believe, especially given the recent interest in agnostic boosting, the current work presents a substantial and concrete improvement over the state of the art, and is a first step in making agnostic boosting practical.

Organization of the appendix. In Appendix A, we point out an error in the branching criteria in [KK09]. Next, in Appendix B, we give a second guarantee on our algorithm that improves the oracle complexity at a vanishing (in ε) cost to the sample complexity. In Appendices C and D, we prove the main result and the lemmas leading up to it. Appendix E provides a proof of the improved result in Appendix B. Appendix F furnishes a proof for the claims concerning extensions to infinite classes. In Appendices G and H, we define the reinforcement learning setup, formally state the RL boosting result along with accompanying algorithms, and prove it. Appendix I substantiates our improved sample complexity bound for learning halfspaces. Finally, in Appendix J, we provide additional experimental details, and Appendix K provides a practical guide for adapting the proposed algorithm.

A Branching criteria in [KK09]

At a high level, the boosting algorithm in [KK09] is an iterative one, that adds either the weak hypothesis or the negation of the sign of the present ensemble to the ensemble mixture in each of the $T \approx 1/\varepsilon^2$ rounds. The algorithm makes use of samples to fulfill two objectives: (a) provide $m \approx 1/\varepsilon^2$ samples to the weak learner to obtain a weak hypothesis that generalizes, (b) use $s \approx 1/\varepsilon^2$ samples to decide whether to add the weak hypothesis or the voting classifier to the current mixture, by comparing the empirical performance on both on said samples. The algorithm uses fresh samples for part (a) every round; this is sound and contributes $Tm \approx 1/\varepsilon^4$ to the net sample complexity.

However, as described in [KK09], the algorithm reuses the same s samples for part (b) across all rounds. This means these s samples determine which of the two choices gets added to the mixture at the end of the first step, and hence H_1. In the next step, however, because H_1 through relabelling of new samples determines the weak hypothesis, these samples are no longer IID with respect to the weak hypothesis or $-\operatorname{sign}(H_1)$, since they have already played a demonstrable part in determining it. This effects occur and compound at all time steps, not just the first. In the analysis, on top of page 7, the analysis in [KK09] uses a Chernoff-Hoeffding bound to say that the performance of the weak hypothesis and $-\operatorname{sign}(H_1)$ on these s samples transfers approximately to the population. However, this inequality may only be applied to IID random variables.

In the case of [KK09], there is a simple and satisfactory fix: resample these s examples from the population distribution every round. The sample complexity, now $T(m+s)$ instead of $Tm+s$, remains $\mathcal{O}(1/\varepsilon^4)$, and no change to the performance gurantee needs to be made.

In our adaptation, this highlights an additional challenge. Even if one were to sate the weak learner with a total fewer number of samples, i.e., perform part (a) with fewer samples, through a uniform convergence result on the base hypothesis (crucially, not the *boosted hypothesis*, whose complexity grows with number of iterations), part (b) requires s fresh samples, and hence $1/\varepsilon^4$ samples in total, in determining which of the weak hypothesis or $-$sign of the ensemble provides a greater magnitude of descent. Here, note that $-\operatorname{sign}(H_t)$ lies outside the base class. To circumvent this, we give a branching criteria to decide which component to add to the present mixture, based on the performance of the weak hypothesis, the only component whose performance is estimable with bounded generalization error, alone on the reused data. Concretely, we introduce a threshold τ to choose the weak hypothesis, if it makes at least τ progress, and otherwise, choose $-\operatorname{sign}(H_t)$ even if in truth it is worse than the weak hypothesis. This deviation requires careful handling in the proof.

B Improved oracle complexity

Here, we provide a different result where Algorithm 1 makes $\mathcal{O}(1/\gamma^2\varepsilon^2)$ call to weak learner, matching exatcly the oracle complexity of existing results, while drawing $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^3 +$

$(\log|\mathcal{B}|)^3/\gamma^2\varepsilon^2)$ samples. Notice that the second term in the sample complexity has a smaller order, whenever ε is sub-constant. The proof may be found in Appendix E.

Theorem 9 (Improved oracle complexity for finite hypotheses class)**.** *Choose any $\varepsilon, \delta > 0$. There exists a choice of $\eta, \sigma, T, \tau, S_0, S, m$ satisfying $T = \mathcal{O}(1/\gamma^2\varepsilon^2), \eta = \mathcal{O}(\gamma^2\varepsilon), \sigma = \eta/\gamma, \tau = \mathcal{O}(\gamma\varepsilon), S = \mathcal{O}((\log|\mathcal{B}|)/\gamma\varepsilon + (\log|\mathcal{B}|)^3), S_0 = \mathcal{O}(1/\varepsilon^2), m = m(\varepsilon_0, \delta_0) + \mathcal{O}(1/\gamma^2\varepsilon^2)$ such that for any γ-agnostic weak learning oracle (as defined in Definition 1) with fixed tolerance ε_0 and failure probability δ_0, Algorithm 1 when run with the potential defined in (1) produces a hypothesis $\overline{h}$ such that with probability $1 - 10\delta_0 T - 10\delta T$,*

$$\mathrm{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}}(h) - \frac{2\varepsilon_0}{\gamma} - \varepsilon,$$

while making $T = \mathcal{O}(1/\gamma^2\varepsilon^2)$ calls to the weak learning oracle, and sampling $TS + S_0 = \mathcal{O}((\log|\mathcal{B}|)/(\gamma^3\varepsilon^3) + (\log|\mathcal{B}|)^3/(\gamma^2\varepsilon^2))$ labeled examples from $\mathcal{D}$.

C Proof of auxiliary lemmas

Proof of Lemma 2. Part 2 is evident from the definition. Taking care that right and left first (and second) derivatives exist and are equal at $z = 0$, by explicit computation, we have

$$\phi'(z) = \begin{cases} -1 & \text{if } z \leq 0, \\ -(z+1)e^{-z} & \text{if } z > 0, \end{cases} \qquad \phi''(z) = \begin{cases} 0 & \text{if } z \leq 0, \\ ze^{-z} & \text{if } z > 0. \end{cases}$$

From this, one may immediately verify Part 1, since ϕ'' is non-negative. Non-negativity of ϕ'' also implies that ϕ' is non-decreasing, and hence, being clearly non-positive, is between $[-1, 0]$; this is Part 3. By elementary calculus, ϕ'' is maximize at $z = 1$ where it equals $1/e$ implying the conclusion in Part 4. □

Proof of Lemma 3. By the definition of $\Phi'_{\mathcal{D}}$, we have

$$\begin{aligned}\Phi'_{\mathcal{D}}(H, \mathrm{sign}(H)) - \Phi'_{\mathcal{D}}(H, h^*) =& \mathbb{E}_{(x,y)\sim\mathcal{D}}[\phi'(yH(x))y(\mathrm{sign}(H)(x) - h^*(x))] \\ =& \mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{1}(yH(x) \geq 0)\phi'(yH(x))y(\mathrm{sign}(H)(x) - h^*(x))] \\ &+ \mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{1}(yH(x) < 0)\phi'(yH(x))y(\mathrm{sign}(H)(x) - h^*(x))].\end{aligned}$$

Note that whenever $yH(x) < 0$, $\phi'(yH(x)) = -1$, which Lemma 2.III attests to. On the other hand, if $yH(x) \geq 0$, since h^* is a binary classifier, $y\mathrm{sign}(H)(x) = 1 \geq yh^*(x)$, and since, again by Lemma 2.III, $\phi'(yH(x)) \geq -1$, we have in this case

$$\phi'(yH(x))y(\mathrm{sign}(H)(x) - h^*(x)) \geq -y(\mathrm{sign}(H)(x) - h^*(x))$$

Plugging these into the previous derivation, we arrive at

$$\begin{aligned}\Phi'_{\mathcal{D}}(H, \mathrm{sign}(H)) - \Phi'_{\mathcal{D}}(H, h^*) \geq& \mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{1}(yH(x) \geq 0)y(h^*(x) - \mathrm{sign}(H)(x))] \\ &+ \mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{1}(yH(x) < 0)y(h^*(x) - \mathrm{sign}(H)(x))] \\ =& \mathbb{E}_{(x,y)\sim\mathcal{D}}[y(h^*(x) - \mathrm{sign}H(x))] \\ =& \mathrm{corr}_{\mathcal{D}}(h^*) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}(H)),\end{aligned}$$

finishing the proof of the claim. □

D Proofs for the main result

Proof of Theorem 4. Let $h^* \in \arg\min_{h\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}}(h)$. Using the update rule for H_{t+1} and the fact that ϕ is 1-smooth (Lemma 2.IV), we arrive at

$$\begin{aligned}\Phi_{\mathcal{D}}(H_{t+1}) &= \mathbb{E}_{(x,y)\sim\mathcal{D}}[\phi(y(H_t(x) + \eta h_t(x)))] \\ &\leq \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi(yH_t(x)) + \eta y\phi'(yH_t(x))h_t(x) + (\eta h_t(x)y)^2/2\right] \\ &\leq \Phi_{\mathcal{D}}(H_t) + \eta\Phi'_{\mathcal{D}}(H_t, h_t) + \eta^2/2\gamma^2,\end{aligned}$$

https://doi.org/10.52202/079017-3190

using the definition of $\Phi'_{\mathcal{D}}$, and that h_t is either a binary classifier, or a $1/\gamma$-scaled version of it. Rearranging this to telescope the sum produces

$$-\frac{1}{T}\sum_{t=1}^{T}\Phi'_{\mathcal{D}}(H_t,h_t) \leq \frac{\sum_{t=1}^{T}(\Phi_{\mathcal{D}}(H_t)-\Phi_{\mathcal{D}}(H_{t+1}))}{\eta T}+\frac{\eta}{2\gamma^2}$$
$$\leq \frac{2}{\eta T}+\frac{\eta}{2\gamma^2} \tag{2}$$

where we use the fact that $\Phi_{\mathcal{D}}(\mathbf{0})=2$, and that it is non-negative (Lemma 2.II).

Hence, $\exists t \in [T]$, such that $-\Phi'_{\mathcal{D}}(H_t,h_t)$ is small. Our proof strategy going forward is to use this fact to imply that both the terms on left side of inequality in Lemma 3, namely, $-\Phi'_{\mathcal{D}}(-\text{sign}(H_t))$ and $-\Phi'_{\mathcal{D}}(H_t,h^*)$, are small for some t, implying a small correlation gap on the population distribution.

Note that if $\text{corr}_{D'_t}(W_t)\geq\tau$, the algorithm sets $h_t=W_t/\gamma$, else it chooses $h_t=-\text{sign}(H_t)$. We analyze these cases separately. For both, Lemma 6 which relates the empirical correlation on $\mathcal{D}_t$ with $\Phi'_{\mathcal{D}}$ will prove indispensable. Going forward, we will define ε_{Gen} as indicated above to capture the generalization error over the base hypothesis class.

For brevity of notation, we condition the analysis going forward on three events. Let $\mathcal{E}_A$ be that event that for all t, $|\text{corr}_{\mathcal{D}_t}(W_t)-\text{corr}_{\widehat{D}'_t}(W_t)|\leq\varepsilon'/10$. Since $\widehat{D}'_t$ is constructed from IID samples from $\mathcal{D}_t$, setting $m\geq 100/\varepsilon'^2\sqrt{\log T/\delta}$, $\Pr(\mathcal{E}_A)\geq 1-\delta$, by an application of Hoeffding's inequality and union bound over t. Here, note that unlike S setting a higher value of m doesn't increase the number of points sampled from $\mathcal{D}$, since D'_t is resampled from already collected data. Similarly, denote by $\mathcal{E}_B$ the event that for all t, $|\text{corr}_{\mathcal{D}}(\text{sign}(H_t))-\text{corr}_{\widehat{D}_0}(\text{sign}(H_t))|\leq\varepsilon''/10$. Again, by Hoeffding's inequality and union bound over t, choosing $S_0=100/\varepsilon''^2\sqrt{\log T/\delta}$, we have $\Pr(\mathcal{E}_B)\geq 1-\delta$, since the samples in $\widehat{D}_0$ were chosen independently of those used to compute H_t's. Finally, we will take the success of Lemma 6 (call this $\mathcal{E}_C$) for granted in the analysis below, for brevity of notation, conditioning our analysis on all three events.

Case A: When $h_t=W_t/\gamma$. When this happens, $\text{corr}_{\mathcal{D}_t}(W_t)\geq\text{corr}_{D'_t}(W_t)-\varepsilon'/10\geq\tau-\varepsilon'/10$. Applying Lemma 6 and noting that $\Phi'_{\mathcal{D}}(H_t,\cdot)$ is linear in its argument, we have

$$-\Phi'_{\mathcal{D}}\left(H_t,\frac{W_t}{\gamma}\right)=-\frac{\Phi'_{\mathcal{D}}(H_t,W_t)}{\gamma}$$
$$\geq\frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\text{corr}_{\mathcal{D}_t}(W_t)-\frac{\varepsilon_{\text{Gen}}}{\gamma}$$
$$\geq\frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau-\frac{\varepsilon'}{10}\right)-\frac{\varepsilon_{\text{Gen}}}{\gamma}$$

Rearranging this

$$\Phi'_{\mathcal{D}}\left(H_t,\frac{W_t}{\gamma}\right)\leq-\frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau-\frac{\varepsilon'}{10}\right)+\frac{\varepsilon_{\text{Gen}}}{\gamma}. \tag{3}$$

Case B: When $h_t=-\text{sign}(H_t)$. Here $\text{corr}_{\mathcal{D}_t}(W_t)\leq\text{corr}_{D'_t}(W_t)+\varepsilon'/10\leq\tau+\varepsilon'/10$. Applying Lemma 6, and using the weak learning condition (Definition 1), we have that

$$\left(1+\frac{\eta}{\sigma}\right)\left(\tau+\frac{\varepsilon'}{10}\right)\geq\left(1+\frac{\eta}{\sigma}\right)\text{corr}_{\mathcal{D}_t}(W_t)$$
$$\geq\gamma\left(1+\frac{\eta}{\sigma}\right)\text{corr}_{\mathcal{D}_t}(h^*)-\left(1+\frac{\eta}{\sigma}\right)\varepsilon_0$$
$$\geq-\gamma\Phi'_{\mathcal{D}}(H_t,h^*)-\left(1+\frac{\eta}{\sigma}\right)\varepsilon_0-\gamma\varepsilon_{\text{Gen}}$$

Now, we invoke the linearity of $\Phi'_{\mathcal{D}}(H_t,\cdot)$ and Lemma 3 to observe that

$$\Phi'_{\mathcal{D}}(H_t,-\text{sign}(H_t))=-\Phi'_{\mathcal{D}}(H_t,\text{sign}(H_t))$$
$$\leq-\Phi'_{\mathcal{D}}(H_t,h^*)-(\text{corr}_{\mathcal{D}}(h^*)-\text{corr}_{\mathcal{D}}(\text{sign}(H_t)))$$
$$\leq-(\text{corr}_{\mathcal{D}}(h^*)-\text{corr}_{\mathcal{D}}(\text{sign}(H_t)))+\frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau+\frac{\varepsilon'}{10}+\varepsilon_0\right)+\varepsilon_{\text{Gen}}. \tag{4}$$

Combining the two. In either case, combining Equations (3) and (4), we have

$$\Phi'_{\mathcal{D}}(H_t, h_t) \leq \max\Bigg\{ -\frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau - \frac{\varepsilon'}{10}\right) + \frac{\varepsilon_{\text{Gen}}}{\gamma}, \\ -\left(\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\text{sign}(H_t))\right) + \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau + \frac{\varepsilon}{10} + \varepsilon_0\right) + \varepsilon_{\text{Gen}}\Bigg\},$$

or using the identity $-\max(-f(x)) = \min f(x)$,

$$-\Phi'_{\mathcal{D}}(H_t, h_t) \geq \min\Bigg\{ \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau - \frac{\varepsilon'}{10}\right) - \frac{\varepsilon_{\text{Gen}}}{\gamma}, \\ \left(\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\text{sign}(H_t))\right) - \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau + \frac{\varepsilon'}{10} + \varepsilon_0\right) - \varepsilon_{\text{Gen}}\Bigg\}.$$

Now, set

$$\tau = \left(1+\frac{\eta}{\sigma}\right)^{-1}\left(\frac{4}{\eta T} + \frac{\eta}{\gamma^2} + \frac{\varepsilon_{\text{Gen}}}{\gamma}\right)\gamma + \frac{\varepsilon'}{10}.$$

Now, either there exists some t such

$$\begin{aligned}\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\text{sign}(H_t)) &\leq \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)(2\tau + \varepsilon_0) + \left(1-\frac{1}{\gamma}\right)\varepsilon_{\text{Gen}} \\ &= \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \left(1+\frac{1}{\gamma}\right)\varepsilon_{\text{Gen}} + \left(1+\frac{\eta}{\sigma}\right)\left(\frac{\varepsilon_0}{\gamma} + \frac{\varepsilon'}{5\gamma}\right), \quad (5)\end{aligned}$$

or the minimum operator in the last expression always accepts the first clause, in which case for all t,

$$-\Phi'_{\mathcal{D}}(H_t, h_t) \geq \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\tau - \frac{\varepsilon'}{10}\right) - \frac{\varepsilon_{\text{Gen}}}{\gamma} = \frac{4}{(\eta T)} + \frac{\eta}{\gamma^2},$$

which contradicts Equation (2). Henceforth let t^* be the iteration for which Equation (5) holds. Finally, given the event $\mathcal{E}_B$, we have

$$\text{corr}_{\mathcal{D}}(\bar{h}) \geq \text{corr}_{\widehat{D}_0}(\bar{h}) - \frac{\varepsilon''}{10} \geq \text{corr}_{\widehat{D}_0}(\text{sign}(H_{t^*})) - \frac{\varepsilon''}{10} \geq \text{corr}_{\mathcal{D}}(\text{sign}(H_{t^*})) - \frac{\varepsilon''}{5}.$$

Compiling this with the inequality in Equation (5), we get

$$\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\bar{h}) \leq \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} + \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\varepsilon_0 + \frac{\varepsilon'}{5}\right) + \frac{\varepsilon''}{5}.$$

Setting $\varepsilon' = \varepsilon/5\gamma$, $\varepsilon'' = \varepsilon/5$ and plugging in the proposed hyper-parameters with appropriate constants yields the claimed result. □

Proof of Lemma 6. Fix any hypothesis $h \in \mathcal{B}$. For any $\eta' \geq 0$, define $H_t^{\eta'} = H_t + \eta' h_t$, and

$$\Delta_t = \Phi'_{\mathcal{D}}(H_t, h) + \left(1+\frac{\eta}{\sigma}\right)\mathbb{E}_{(x,y)\sim\mathcal{D}_t}\left[yh(x)\right].$$

First, we derive recursive expansions of $\Phi'_{\mathcal{D}}(H_t, h)$ and $\mathcal{D}_t$, the two quantities we wish to relate.

Claim 10. *For any t and $h : \mathcal{X} \to \mathbb{R}$, we have*

$$\Phi'_{\mathcal{D}}(H_t, h) = \Phi'_{\mathcal{D}}(H_{t-1}, h) + \eta\mathbb{E}_{\eta'\sim Unif[0,\eta]}[\Phi''_{\mathcal{D}}(H_{t-1}^{\eta'}, h, h_{t-1})].$$

Claim 11. *For any t,*

$$\mathbb{E}_{t-1}[\mathbb{E}_{(x,y)\sim\mathcal{D}_t}\left[yh(x)\right]] \\ = (1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}\left[yh(x)\right] - \frac{\sigma^2}{\sigma+\eta}\Phi'_{\mathcal{D}}(H_{t-1}, h) - \frac{\eta\sigma}{\sigma+\eta}\mathbb{E}_{\eta'\sim Unif[0,\eta]}[\Phi''_{\mathcal{D}}(H_{t-1}^{\eta'}, h, h_{t-1})].$$

https://doi.org/10.52202/079017-3190

Adding $(1+\eta/\sigma)$ times the last expression to the expansion of $\Phi'_{\mathcal{D}}(H_t, h)$, we have

$$\begin{aligned}\mathbb{E}_{t-1}[\Delta_t] &= \Phi'_{\mathcal{D}}(H_t, h) + \left(1+\frac{\eta}{\sigma}\right)\mathbb{E}_{t-1}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}_t}\left[yh(x)\right]\right] \\ &= (1-\sigma)\Phi'_{\mathcal{D}}(H_{t-1}, h) + (1-\sigma)\left(1+\frac{\eta}{\sigma}\right)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}\left[yh(x)\right] \\ &= (1-\sigma)\Delta_{t-1}.\end{aligned}$$

Using this, we conclude that $\Delta'_t = (1-\sigma)^{-t}\Delta_t$ forms a martingale sequence with respect to the $\{\mathcal{F}_t : t \in \mathbb{N}_{\geq 0}\}$ filtration sequence, as $\Delta'_t = \mathbb{E}_{t-1}[(1-\sigma)^{-t}\Delta_t] = (1-\sigma)^{t-1}\Delta_{t-1} = \Delta'_{t-1}$. The associated martingale difference sequence $\delta'_t = \Delta'_t - \Delta'_{t-1}$ can be bounded both in worst-case and second-moment terms as we show next.

Claim 12. *For all t, $|\delta'_t| \leq (1-\sigma)^{-t}2(\sigma+\eta/\gamma)$ and $\sum_{s=1}^{t}\mathbb{E}_{s-1}[{\delta'_s}^2] \leq \frac{8(\sigma^2+\eta^2/\gamma^2)}{\sigma(1-\sigma)^{2t}S}$.*

Now, we are ready to apply Freedman's inequality for martingales.

Theorem 13 (Freedman's inequality [Fre75])**.** *Consider a real-valued martingale $\{Y_k : k \in \mathbb{Z}_{\geq 0}\}$ with respect to some filtration sequence $\{\Sigma_k : k \in \mathbb{Z}_{\geq 0}\}$, and let $\{X_k : k \in \mathbb{Z}_{>0}\}$ be the associated difference sequence. Assume that the difference sequence is uniformly bounded: $|X_k| \leq R$ almost surely for $k \in \mathbb{Z}_{>0}$. Define the predictable quadratic variation process: $W_k := \sum_{j=1}^{k}\mathbb{E}_{j-1}(X_j^2)$ for $k \in \mathbb{Z}_{>0}$, where $\mathbb{E}_{j-1}(\,\cdot\,) := \mathbb{E}(\,\cdot\,|\Sigma_{j-1})$. Then, for all $t \geq 0$ and $\sigma^2 > 0$,*

$$\Pr\left(\exists k \geq 0 : Y_k \geq t \;\text{ and }\; W_k \leq \sigma^2\right) \leq \exp\left\{-\frac{-t^2/2}{\sigma^2 + Rt/3}\right\}.$$

Applying Freedman's inequality to Δ'_t, since the total conditional variance of our martingale difference sequence is bounded almost surely as shown before, we get for any

$$r \geq 4(1-\sigma)^{-t}\left(\sigma+\frac{\eta}{\gamma}\right)\max\left\{\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{2}{\delta}}, \log\frac{2}{\delta}\right\}$$

that

$$\Pr(\Delta'_t \geq r) \leq \exp\left\{-\left(\frac{r^2}{2}\right)\Big/\left(\frac{8(\sigma^2+\eta^2/\gamma^2)}{\sigma(1-\sigma)^{2t}S} + \frac{2(\sigma+\eta/\gamma)r}{(1-\sigma)^t}\right)\right\} \leq \delta.$$

Hence, using $\Delta'_t = (1-\sigma)^{-t}\Delta_t$, with probability $1-\delta$, we have

$$\Delta_t \leq 4\left(\sigma+\frac{\eta}{\gamma}\right)\left(\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{2}{\delta}} + \log\frac{2}{\delta}\right).$$

Taking a union bound over the choice of h from $\mathcal{B}\cup\{h^*\}$ and over t, along with repeating the argument on the martingale $-\Delta_t$ to furnish the promised two-sided bound, concludes the claim. □

Proof of Claim 10. Using the fundamental theorem of calculus and that $\phi \in \mathcal{C}^2$, which Lemma 2.I certifies, we get

$$\begin{aligned}&\Phi'_{\mathcal{D}}(H_t, h) \\ &= \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi'(yH_t(x))yh(x)\right] \\ &= \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi'(y(H_{t-1}(x)+\eta h_{t-1}(x)))yh(x)\right] \\ &= \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi'(yH_{t-1}(x))yh(x) + \int_0^{\eta}\phi''(y(H_{t-1}(x)+\eta' h_{t-1}(x)))y^2h_{t-1}(x)h(x)d\eta'\right] \\ &= \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi'(yH_{t-1}(x))yh(x)\right] + \eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}\mathbb{E}_{(x,y)\sim\mathcal{D}}[h(x)h_{t-1}(x)\phi''(yH_{t-1}^{\eta'}(x))] \\ &= \Phi'_{\mathcal{D}}(H_{t-1}, h) + \eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H_{t-1}^{\eta'}, h, h_{t-1})].\end{aligned}$$

where use the fact that for binary labels $y \in \{-1, 1\}$, $y^2 = 1$, and the definitions of $\Phi'_{\mathcal{D}}$ and $\Phi''_{\mathcal{D}}$. □

Proof of Claim 11. First, we establish a recursive structure on the random random distribution $\mathcal{D}_t$ without any conditional expectation in place.

Claim 14. *For any t, we have*

$$\begin{aligned}\mathbb{E}_{(x,y)\sim\mathcal{D}_t}[yh(x)] &= (1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)] \\ &\quad - \frac{\sigma}{\sigma+\eta}\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\sigma\phi'(yH_{t-1}(x))yh(x) + \eta\mathbb{E}_{\eta'\sim Unif[0,\eta]}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]\right].\end{aligned}$$

Using the fact that $\widehat{D}_t$ identically samples data points from $\mathcal{D}$, we arrive at

$$\begin{aligned}&\mathbb{E}_{t-1}[\mathbb{E}_{(x,y)\sim\mathcal{D}_t}[yh(x)]] \\ =&(1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)] \\ &- \frac{\sigma^2}{\sigma+\eta}\mathbb{E}_{(x,y)\sim\mathcal{D}}[\phi'(yH_{t-1}(x))yh(x)] + \frac{\eta\sigma}{\sigma+\eta}\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)\right] \\ =&(1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)] - \frac{\sigma^2}{\sigma+\eta}\Phi'_{\mathcal{D}}(H_{t-1},h) - \frac{\eta\sigma}{\sigma+\eta}\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1})].\end{aligned}$$

□

Proof of Claim 14. By the definition of $\mathcal{D}_t$, we have

$$\begin{aligned}&\mathbb{E}_{(x,y)\sim\mathcal{D}_t}[yh(x)] \\ =&(1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)] \\ &- \sigma\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[h(x)\left(\frac{\sigma}{\sigma+\eta}\phi'(yH_{t-1}(x))y + \frac{\eta}{\eta+\sigma}\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)]\right)\right] \\ =&(1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)] \\ &- \frac{\sigma}{\sigma+\eta}\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\sigma\phi'(yH_{t-1}(x))yh(x) + \eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]\right]\end{aligned}$$

where in the first equality we use the fact that for any binary random variable Y supported on $\{-1,1\}$ with $\Pr(Y=1)=p$, we have $\mathbb{E}[Y]=2p-1$. □

Proof of Claim 12. Using the recursive expansion of $\Phi'_{\mathcal{D}}$ (Claim 10) and $\mathcal{D}_t$ (Claim 14), we have

$$\begin{aligned}\delta'_t =&(1-\sigma)^{-t}\left(\Phi'_{\mathcal{D}}(H_t,h) - (1-\sigma)\Phi'_{\mathcal{D}}(H_{t-1},h)\right) \\ &+ (1-\sigma)^{-t}\left(1+\frac{\eta}{\sigma}\right)\left(\mathbb{E}_{(x,y)\sim\mathcal{D}_t}[yh(x)] - (1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)]\right) \\ =&(1-\sigma)^{-t}\left(\sigma\Phi'_{\mathcal{D}}(H_{t-1},h) + \eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1})]\right) \\ &- (1-\sigma)^{-t}\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\sigma\phi'(yH_{t-1}(x))yh(x) + \eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]\right] \\ =&(1-\sigma)^{-t}\sigma\left(\Phi'_{\mathcal{D}}(H_{t-1},h) - \mathbb{E}_{(x,y)\sim\widehat{D}_t}[\phi'(yH_{t-1}(x))yh(x)]\right) \\ &+ (1-\sigma)^{-t}\eta\left(\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1}) - \mathbb{E}_{(x,y)\sim\widehat{D}_t}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]]\right).\end{aligned}$$

Using Lemma 2, ϕ', ϕ'' are uniformly bounded in magnitude by one, $\Phi'_{\mathcal{D}}(H,\cdot)$ and $\Phi''_{\mathcal{D}}(H,\cdot,g)$ are uniformly bounded in magnitude by one for any $H:\mathcal{X}\to\mathbb{R}$ and $1/\gamma$-uniformly bounded function g. Hence, $|\delta'_t| \le (1-\sigma)^{-t}2(\sigma+\eta/\gamma)$. To bound the conditional variance, we use the identity

https://doi.org/10.52202/079017-3190

$(a+b)^2 \le 2a^2 + 2b^2$ in the first line below to show

$$
\begin{aligned}
&\mathbb{E}_{t-1}[{\delta'_t}^2] \\
\le& 2(1-\sigma)^{-2t}\Bigg(\sigma^2\mathbb{E}_{t-1}\left[\Phi'_{\mathcal{D}}(H_{t-1},h) - \mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\phi'(yH_{t-1}(x))yh(x)\right]\right]^2 \\
&+ \eta^2\mathbb{E}_{t-1}\left[\mathbb{E}_{\eta'\sim\mathrm{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1}) - \mathbb{E}_{(x,y)\sim\widehat{D}_t}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]]\right]^2\Bigg) \\
=& 2(1-\sigma)^{-2t}S^{-1}\Bigg(\sigma^2\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\Phi'_{\mathcal{D}}(H_{t-1},h) - \phi'(yH_{t-1}(x))yh(x)\right]^2 \\
&+ \eta^2\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\mathbb{E}_{\eta'\sim\mathrm{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1}) - \phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]\right]^2\Bigg) \\
\le& \frac{8(\sigma^2+\eta^2/\gamma^2)}{(1-\sigma)^{2t}S},
\end{aligned}
$$

where we use the fact that the S samples consisting $\widehat{D}_t$ are independent and identically sampled from $\mathcal{D}$ conditioned on $\mathcal{F}_{t-1}$ in the second equality, and that the expectation of any function over $\widehat{D}_t$ is the equivalent sample average. Finally, using an identity on geometric sums, we get

$$
\sum_{s=1}^{t}\mathbb{E}_{s-1}[{\delta'_s}^2] \le \frac{8(\sigma^2+\eta^2/\gamma^2)}{S}\frac{(1-\sigma)^{-2(t+1)}-1}{(1-\sigma)^{-2}-1} \le \frac{8(\sigma^2+\eta^2/\gamma^2)}{\sigma(1-\sigma)^{2t}S},
$$

where we appeal to the inequality $1-(1-\sigma)^2 = 2\sigma-\sigma^2 \ge \sigma$ for any $\sigma \in [0,1]$. □

E Proofs for improved oracle complexity

Proof of Theorem 9. The overall proof is similar to that of Theorem 4, with the exception of the following bound on ε_{Gen}, which we state now (and prove next) in a new lemma relating the empirical correlation on $\mathcal{D}_t$ with $\Phi'_{\mathcal{D}}$. The principal difference between Lemma 6 and Lemma 15 is the presence of a $1/\sqrt{S}$ in both terms on the right in Lemma 15, however this comes at the cost of a higher polynomial dependence in $\log|\mathcal{B}|$ in the second term on the right side of Lemma 15.

Lemma 15. *There exists a $C>0$ such that with probability $1-\delta$, for all $t\in[T]$ and $h\in\mathcal{B}\cup\{h^*\}$, we have*

$$
\left|\Phi'_{\mathcal{D}}(H_t,h) + \left(1+\frac{\eta}{\sigma}\right)\mathrm{corr}_{\mathcal{D}_t}(h)\right| \le \underbrace{C\left(\sigma+\frac{\eta}{\gamma}\right)\left(\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{|\mathcal{B}|T}{\delta}} + \frac{1}{\sqrt{S}}\left(\log\frac{|\mathcal{B}|T}{\delta}\right)^{3/2}\right)}_{:=\varepsilon_{Gen}}.
$$

From here, with the new definition of ε_{Gen} as defined above, identically following the steps in the proof of Theorem 4, which we skip to avoid repetition, we arrive at

$$
\mathrm{corr}_{\mathcal{D}}(h^*) - \mathrm{corr}_{\mathcal{D}}(\overline{h}) \le \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} + \frac{1}{\gamma}\left(1+\frac{\eta}{\sigma}\right)\left(\varepsilon_0 + \frac{\varepsilon'}{5}\right) + \frac{\varepsilon''}{5}.
$$

Setting $\varepsilon' = \varepsilon/5\gamma$, $\varepsilon'' = \varepsilon/5$ and plugging in the proposed hyper-parameters with appropriate constants yields the claimed result. □

Proof of Lemma 15. The proof of the present lemma is largely similar to that of Lemma 6, with two exceptions: (a) we use a variant of Freedman's inequality that applies when the martingale difference sequences (for us, δ'_t are bounded with high probability, instead of admitting an almost sure absolute bound; (b) to apply this result, we establish a high probability upper bound on δ'_t that scales as $1/\sqrt{S}$.

Fix any hypothesis $h\in\mathcal{B}$. For any $\eta'\ge 0$, define $H^{\eta'}_t = H_t + \eta' h_t$, and

$$
\Delta_t = \Phi'_{\mathcal{D}}(H_t,h) + \left(1+\frac{\eta}{\sigma}\right)\mathbb{E}_{(x,y)\sim\mathcal{D}_t}\left[yh(x)\right].
$$

As before, adding $(1+\eta/\sigma)$ times the expression in Claim 11 to the the expansion of $\Phi'_{\mathcal{D}}(H_t, h)$ in Claim 10, we observe that $\mathbb{E}_{t-1}[\Delta_t] = (1-\sigma)\Delta_{t-1}$, and hence conclude that $\Delta'_t = (1-\sigma)^{-t}\Delta_t$ forms a martingale sequence with respect to the $\{\mathcal{F}_t : t \in \mathbb{N}_{\geq 0}\}$ filtration sequence. As shown in Claim 12, the associated martingale difference sequence $\delta'_t = \Delta'_t - \Delta'_{t-1}$ admits a total variance bound of $\sum_{s=1}^{t} \mathbb{E}_{s-1}[{\delta'_s}^2] \leq \frac{8(\sigma^2+\eta^2/\gamma^2)}{\sigma(1-\sigma)^{2t}S}$.

Now, we state a variant of Freedman's inequality that applies when martingale difference sequences are bounded with high probability.

Theorem 16 (Freedman's inequality with high probability bounds [Pin94, Pin20])**.** *Fix a positive integer n. Consider a real-valued martingale $\{Y_k : k \in \mathbb{Z}_{\geq 0}\}$ with respect to some filtration sequence $\{\Sigma_k : k \in \mathbb{Z}_{\geq 0}\}$, and let $\{X_k : k \in \mathbb{Z}_{>0}\}$ be the associated difference sequence. Assume that the difference sequence is uniformly bounded with high probability for some R, δ':*

$$\Pr\left(\max_{k\in[n]} |X_k| \geq R\right) \leq \delta'.$$

Define the predictable quadratic variation process: $W_n := \sum_{j=1}^{n} \mathbb{E}_{j-1}(X_j^2)$, where $\mathbb{E}_{j-1}(\,\cdot\,) := \mathbb{E}(\,\cdot\,|\Sigma_{j-1})$. Then, for all $t \geq 0$ and $\sigma^2 > 0$,

$$\Pr\left(Y_n \geq t \text{ and } W_n \leq \sigma^2\right) \leq \exp\left\{-\frac{-t^2/2}{\sigma^2 + Rt/3}\right\} + \delta'.$$

To apply this variant of Freedman's inequality, we establish a high probability bound on Δ'_t, that is, ignoring logarithmic terms, substantially better than the almost sure bound in Claim 12.

Claim 17. *There exists a universal constant C, such that for any t, we have*

$$\Pr\left(\max_{s\in[t]} |\delta'_s| \geq \frac{C}{(1-\sigma)^t}\left(\sigma + \frac{\eta}{\gamma}\right)\frac{1}{\sqrt{S}}\sqrt{\log\frac{t}{\delta}}\right) \leq \delta.$$

Combining this with the almost sure bound on the total conditional variance of our martingale difference sequence, we get for any

$$r \geq 4(1-\sigma)^{-t}\left(\sigma + \frac{\eta}{\gamma}\right)\max\left\{\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{2}{\delta}}, \frac{C}{\sqrt{S}}\left(\log\frac{2t}{\delta}\right)^{3/2}\right\}$$

that

$$\Pr(\Delta'_t \geq r) \leq \exp\left\{-\left(\frac{r^2}{2}\right) \Big/ \left(\frac{8(\sigma^2+\eta^2/\gamma^2)}{\sigma(1-\sigma)^{2t}S} + \frac{C(\sigma+\eta/\gamma)r}{(1-\sigma)^t\sqrt{S}}\sqrt{\log\frac{t}{\delta}}\right)\right\} + \delta \leq 2\delta.$$

Hence, using $\Delta'_t = (1-\sigma)^{-t}\Delta_t$, with probability $1-2\delta$, we have

$$\Delta_t \leq 4\left(\sigma + \frac{\eta}{\gamma}\right)\left(\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{2}{\delta}} + \frac{C}{\sqrt{S}}\left(\log\frac{2t}{\delta}\right)^{3/2}\right).$$

Taking a union bound over the choice of h from $\mathcal{B} \cup \{h^*\}$ and over t, along with repeating the argument on the martingale $-\Delta_t$ to furnish the promised two-sided bound, concludes the claim. □

https://doi.org/10.52202/079017-3190

Proof of Claim 17. The proof begins similarly to the one for Claim 12. Using the recursive expansion of $\Phi'_{\mathcal{D}}$ (Claim 10) and $\mathcal{D}_t$ (Claim 14), we have

$$\begin{aligned}
\delta'_t =&(1-\sigma)^{-t}\left(\Phi'_{\mathcal{D}}(H_t,h)-(1-\sigma)\Phi'_{\mathcal{D}}(H_{t-1},h)\right)\\
&+(1-\sigma)^{-t}\left(1+\frac{\eta}{\sigma}\right)\left(\mathbb{E}_{(x,y)\sim\mathcal{D}_t}[yh(x)]-(1-\sigma)\mathbb{E}_{(x,y)\sim\mathcal{D}_{t-1}}[yh(x)]\right)\\
=&(1-\sigma)^{-t}\left(\sigma\Phi'_{\mathcal{D}}(H_{t-1},h)+\eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1})]\right)\\
&-(1-\sigma)^{-t}\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\sigma\phi'(yH_{t-1}(x))yh(x)+\eta\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]\right]\\
=&(1-\sigma)^{-t}\sigma\left(\underbrace{\Phi'_{\mathcal{D}}(H_{t-1},h)-\mathbb{E}_{(x,y)\sim\widehat{D}_t}\left[\phi'(yH_{t-1}(x))yh(x)\right]}_{:=A_1}\right)\\
&+(1-\sigma)^{-t}\eta\left(\mathbb{E}_{\eta'\sim\text{Unif}[0,\eta]}[\underbrace{\Phi''_{\mathcal{D}}(H^{\eta'}_{t-1},h,h_{t-1})-\mathbb{E}_{(x,y)\sim\widehat{D}_t}[\phi''(yH^{\eta'}_{t-1}(x))h_{t-1}(x)h(x)]}_{:=A_2}]\right).
\end{aligned}$$

However, this time, we note that, conditioned on $\mathcal{F}_{t-1}$, A_1 and A_2 are averages of S zero-mean IID random variables, where each constituent is absolutely bounded by one in magnitude (Lemma 2). Hence, by Hoeffding's inequality, with $1-2\delta$, we have that

$$\max\{|A_1|,|A_2|\}\leq\frac{100}{\sqrt{S}}\sqrt{\log\frac{2}{\delta}}.$$

Note that since this statement holds true for all realizations in $\mathcal{F}_{t-1}$ – in words, it is a statement about the randomness in the present round alone – it remains true when not subject to such filtration's, i.e., by marginalizing over $\mathcal{F}_{t-1}$. A union bound over t now yields the claim □

F Proofs for extensions to infinite classes

Definition 18 (Covering number)**.** Given a set $\mathcal{F}$ in linear space $\mathcal{V}$, a semi norm $\|\cdot\|$ on $\mathcal{V}$, $\mathcal{N}(\varepsilon,\mathcal{F},\|\cdot\|)$ is the size of the smallest set $\mathcal{G}$ such that for all $f\in\mathcal{F}$, there exists a $g\in\mathcal{G}$ such that $\|f-g\|\leq\varepsilon$.

Proof of Theorem 5. First, we state (and subsequently prove) the following sample complexity result using L_1 covering numbers. We note that it may be possible to improve this result for specific classes of functions, i.e., monotonic functions, by applying chaining techniques [Dud74] to L_2 distances.

Theorem 19 (Main result for Covering Number)**.** *Define $L_{1,\mathcal{D}_{\mathcal{X}}}(f,g)=\mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}}[|f(x)-g(x)|]$, for any two functions $f,g:\mathcal{X}\to\mathbb{R}$, where $\mathcal{D}_{\mathcal{X}}$ is the marginal distribution of $\mathcal{D}$ on the features set $\mathcal{X}$. There exists a setting of parameters such that for any for any γ-agnostic weak learning oracle with fixed tolerance ε_0 and failure probability δ_0, Algorithm 1 produces a hypothesis $\overline{h}$ such that with probability $1-10\delta_0T-10\delta T$,*

$$\text{corr}_{\mathcal{D}}(\overline{h})\geq\max_{h\in\mathcal{H}}\text{corr}_{\mathcal{D}}(h)-\frac{2\varepsilon_0}{\gamma}-\varepsilon,$$

while making $T=\mathcal{O}((\log\mathcal{N}(\varepsilon/10\gamma,\mathcal{B},L_{1,\mathcal{D}_{\mathcal{X}}}))/\gamma^2\varepsilon^2)$ calls to the weak learning oracle, and sampling $TS+S_0=\mathcal{O}((\log\mathcal{N}(\varepsilon/10\gamma,\mathcal{B},L_{1,\mathcal{D}_{\mathcal{X}}}))/\gamma^3\varepsilon^3)$ labeled examples from $\mathcal{D}$.

The claim follows immediately by the application of the following result from [VDVW97] to place an upper bound of the L_1 covering in terms of the VC dimension in Theorem 19. This result originally due to [Hau95] was established for distances defined by n-point empirical measures, for some finite n. The version in [VDVW97] works on arbitrary distributions, by first proving that by a result proven for empirical-type measures transfers without loss to any distribution.

Theorem 20 (Theorem 2.6.4 in [VDVW97])**.** *There exists a universal constant C such that for any VC class $\mathcal{B}$, any probability measure $\mathcal{D}_{\mathcal{X}}$, and $0\leq\varepsilon\leq1$,*

$$\mathcal{N}(\varepsilon,\mathcal{B},L_{1,\mathcal{D}_{\mathcal{X}}})\leq C\cdot VC(\mathcal{B})\cdot\left(\frac{4e}{\varepsilon}\right)^{VC(\mathcal{B})}.$$

This concludes the proof. □

Proof of Theorem 19. We first invoke Lemma 6 but on a minimal ε_{Cov}-cover (say $\mathcal{B}_{\text{Cov}}$) of $\mathcal{B}$ with respect to the $L_{1,\mathcal{D}_\mathcal{X}}$ semi norm to get that there exists a $C > 0$ such that with probability $1-\delta$, for all $t \in [T]$ and $h \in \mathcal{B}_{\text{Cov}} \cup \{h^*\}$, we have

$$\begin{aligned}&\left|\Phi'_\mathcal{D}(H_t, h) + \left(1 + \frac{\eta}{\sigma}\right)\text{corr}_{\mathcal{D}_t}(h)\right| \\ &\leq \underbrace{C\left(\sigma + \frac{\eta}{\gamma}\right)\left(\frac{1}{\sqrt{\sigma S}}\sqrt{\log\frac{T\log\mathcal{N}(\varepsilon_{\text{Cov}}, \mathcal{B}, L_{1,\mathcal{D}_\mathcal{X}})}{\delta}} + \log\frac{T\log\mathcal{N}(\varepsilon_{\text{Cov}}, \mathcal{B}, L_{1,\mathcal{D}_\mathcal{X}})}{\delta}\right)}_{:=\varepsilon'_{\text{Gen}}}.\end{aligned}$$

Fix any $g \in \mathcal{B}$. Now, by virtue of the ε_{Cov}-cover, there exists a $h \in \mathcal{B}$ such that $L_{1,\mathcal{D}_\mathcal{X}} = \mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}}[|f(x) - g(x)|] \leq \varepsilon_{\text{Cov}}$. Now, using the definition of $\Phi'_\mathcal{D}$ and a uniform (absolute) upper bound on ϕ' (Lemma 2.III), we get

$$\begin{aligned}|\Phi'_\mathcal{D}(H_t, h) - \Phi'_\mathcal{D}(H_t, g)| &= |\mathbb{E}_{(x,y)\sim\mathcal{D}}[y\phi'(yH_t(x))(h(x) - g(x))]| \\ &\leq \mathbb{E}_{(x,y)\sim\mathcal{D}}[|y\phi'(yH_t(x))| \cdot |h(x) - g(x)|] \\ &\leq \mathbb{E}_{(x,y)\sim\mathcal{D}}[|h(x) - g(x)|] \\ &\leq \mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}}[|h(x) - g(x)|] \leq \varepsilon_{\text{Cov}}.\end{aligned}$$

Similarly, using the fact that the marginal distribution of $\mathcal{D}_t$ on $\mathcal{X}$ is the same as $\mathcal{D}_\mathcal{X}$, we get

$$|\text{corr}_{\mathcal{D}_t}(h) - \text{corr}_{\mathcal{D}_t}(g)| \leq \mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}}[|h(x) - g(x)|] \leq \varepsilon_{\text{Cov}}.$$

Combining these, we have that with probability $1-\delta$, for all $t \in [T]$ and $h \in \mathcal{B}_{\text{Cov}} \cup \{h^*\}$, we have

$$\left|\Phi'_\mathcal{D}(H_t, h) + \left(1 + \frac{\eta}{\sigma}\right)\text{corr}_{\mathcal{D}_t}(h)\right| \leq \varepsilon'_{\text{Gen}} + 2\varepsilon_{\text{Cov}}.$$

From here, proceeding identically as the steps in the proof of Theorem 4 yields

$$\text{corr}_\mathcal{D}(h^*) - \text{corr}_\mathcal{D}(\overline{h}) \leq \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \frac{2(\varepsilon'_{\text{Gen}} + \varepsilon_{\text{Cov}})}{\gamma} + \frac{1}{\gamma}\left(1 + \frac{\eta}{\sigma}\right)\left(\varepsilon_0 + \frac{\varepsilon'}{5}\right) + \frac{\varepsilon''}{5}.$$

Setting $\varepsilon' = \varepsilon/5\gamma$, $\varepsilon'' = \varepsilon/5$, $\varepsilon_{\text{Cov}} = \varepsilon/10\gamma$ and plugging in the proposed hyper-parameters with appropriate constants yields the claimed result. □

G Boosting for reinforcement learning

MDP. In this section, we consider a Markov Decision Process $\mathcal{M} = (\mathcal{S}, \mathcal{A}, r, P, \beta, \mu_0)$, where $\mathcal{S}$ is a set of states, $\mathcal{A} = \{\pm 1\}$ is a binary set of actions, $r : \mathcal{S} \times \mathcal{A} \to [0, 1]$ determines the (expected) reward at any state-action pair, $P : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ captures the transition dynamics of the MDP, i.e., $P(s'|s, a)$ is the probability of moving to state s' upon taking action a at state s, $\beta \in [0, 1)$ is the discount factor, and μ_0 is the initial state distribution. Without any loss, one may restrict their consideration to Markovian policies [Put14] of the form $\pi : \mathcal{S} \to \Delta(\mathcal{A})$, where an agent at each point in time chooses action a at state s independently with probability $\pi(a|s)$.

For any state $s \in \mathcal{S}$, action $a \in \mathcal{A}$, and distribution $\mu \sim \Delta(\mathcal{S})$ over states, define state-action and state-value functions as

$$\begin{aligned}Q^\pi(s, a) &= \mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t r(s_t, a_t)\middle|\pi, s_0 = s, a_0 = a\right], \\ V^\pi(s) &= \mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t r(s_t, a_t)\middle|\pi, s_0 = s\right], \\ V^\pi_\mu &= \mathbb{E}_{s\sim\mu}\left[V^\pi(s)\right].\end{aligned}$$

https://doi.org/10.52202/079017-3190

Algorithm 2 RL Boosting adapted from [BHS22]

Input: iteration budget T, state distribution μ, step sizes η_t, post-selection sample size P
Initialize a policy $\pi_0 \in \Pi$ arbitrarily.
for $t = 1$ **to** T **do**
    Run Algorithm 1 to get π'_t, while using Algorithm 3 to produce a distribution over state-actions (ignore $\widehat{Q}$) by executing the current policy π_{t-1} starting from the initial state distribution μ.
    Update $\pi_t = (1 - \eta_t)\pi_{t-1} + \eta_t \pi'_t$.
end for
Run each policy π_t for P rollouts to compute an empirical estimate $\widehat{V^{\pi_t}}$ of the expected return.
return $\overline{\pi} = \pi_{t'}$ where $t' = \arg\max_t \widehat{V^{\pi_t}}$.

We will abbreviate $V^{\pi}_{\mu_0} = V^{\pi}$, since it captures the value of any policy starting from the canonical starting state distribution. Finally, the occupancy measure $\mu^{\pi}_{\mu'}$ induced by a policy π starting from an initial state distribution μ' is stated below. We will take $\mu^{\pi} = \mu^{\pi}_{\mu_0}$ as a matter of convention.

Accessing the MDP. Following [BHS22], we consider two models of accessing the MDP; we will furnish a different result for each. In the *episodic model*, the learner interacts with the MDP in a limited number of episodes of reasonable length (i.e., $\approx (1-\beta)^{-1}$), and the starting state of MDP is always drawn from μ_0. In the second, termed *rollouts with ν-resets*, the learner's interaction is still limited to a small number of episodes, however, the MDP now samples its starting state from ν. It is important to stress that in both cases, the learner's objective is the same, to maximize V^{π} starting from μ_0. However, ν could be more *spread out* over the state space than μ_0, and provide an implicit source of explanation, and the learner's guarantee as shown next benefits from its dependence on a milder notion of distribution mismatch in this case.

Weak Leaner. For convenience, we restate our weak learning definition, this time using π to denote policies, instead of h. Note that our definition is equivalent to that used by [BHS22], because for binary actions a random policy induces an accuracy of half regardless of the distribution over features and labels. One may use the identity $\mathrm{corr}_{\mathcal{D}}(\pi) = 1 - 2l_{\mathcal{D}}(\pi)$ to observe this. In fact, our assumption of weak learning might ostensibly seem weaker, since it operates with $0/1$ losses (equivalently, correlations), whereas the losses in previous work are assumed general linear. However, for binary actions, this difference is insubstantial and purely stylistic.

Definition 21 (Agnostic Weak Learner)**.** A learning algorithm is called a γ-agnostic weak learner with sample complexity $m : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{N} \cup \{+\infty\}$ with respect to a policy class Π and a base policy class $\mathcal{B}$ if, for any $\varepsilon_0, \delta_0 > 0$, upon being given $m(\varepsilon_0, \delta_0)$ independently and identically distributed samples from any distribution $\mathcal{D}' \in \Delta(\mathcal{X} \times \{\pm 1\})$, it can output a base policy $\mathcal{W} \in \mathcal{B}$ such that with probability $1 - \delta_0$ we have

$$\mathrm{corr}_{\mathcal{D}'}(\mathcal{W}) \geq \gamma \max_{\pi \in \Pi} \mathrm{corr}_{\mathcal{D}'}(\pi) - \varepsilon_0.$$

Policy Completeness and Distribution Mismatch. $\pi^* \in \arg\max_{\pi} V^{\pi}$ be a reward maximizing policy, and V^* be its value. Let $\mathbb{\Pi}$ be the convex hull of the boosted policy class, i.e., the outputs of the boosting algorithm. For any state distribution μ', define the policy completeness $\mathcal{E}_{\mu'}$ term as

$$\mathcal{E}_{\mu'} = \max_{\pi \in \mathbb{\Pi}} \min_{\pi' \in \Pi} \mathbb{E}_{s \sim \mu^{\pi}_{\mu'}} [\max_{a \in \mathcal{A}} Q^{\pi}(s,a) - \mathbb{E}_{a \sim \pi'(\cdot|s)} Q^{\pi}(s,a)].$$

In words, this term captures how well the greedy policy improvement operator is approximated by Π in an state-averaged sense over the distribution induces by any policy in $\mathbb{\Pi}$. Finally, we define distribution mismatch coefficients below.

$$C_{\infty} = \max_{\pi \in \mathbb{\Pi}} \|\mu^{\pi^*}/\mu^{\pi}\|_{\infty}, \quad D_{\infty} = \|\mu^{\pi^*}/\nu\|_{\infty}.$$

Theorem 22. *Let $\mathcal{W}$ be a γ-weak learner for the policy class Π operating with a base class $\mathcal{B}$, with sample complexity $m(\varepsilon_0, \delta_0) = (\log|\mathcal{B}|/\delta_0)/\varepsilon_0^2$. Fix tolerance ε and failure probability δ. In the* episodic *access model, there is a setting of parameters such that Algorithm 2 when given access to $\mathcal{W}$ produces a policy $\overline{\pi}$ such that with probability $1 - \delta$, we have*

$$V^* - V^{\overline{\pi}} \leq \frac{C_{\infty}\mathcal{E}}{1-\beta} + \varepsilon,$$

Algorithm 3 Trajectory Sampler adapted from [BHS22]

1: Sample state $s_0 \sim \mu$ and action $a' \sim \text{Unif}(\mathcal{A})$.
2: Sample $s \sim \mu^\pi$ as follows: at every step h, with probability β, execute π; else, accept s_h.
3: Take action a' at state s_h, then continue to execute π, and use a termination probability of $1-\beta$. Upon termination, set $R(s_h, a')$ as the sum of rewards from time h onwards.
4: Define the vector $\widehat{Q}$, such that for all $a \in A$, $\widehat{Q}(a) = 2R(s_h, a') \cdot \mathbb{I}_{a=a'}$.
5: With probability $C\widehat{Q}(a')$, set $y = a'$ else set $y \in \mathcal{A} - \{a'\}$, where $C = (1-\beta)/2$.
6: **return** $(s_h, \widehat{Q}, y)$.

while sampling

$$\mathcal{O}\left(\frac{C_\infty^5 \log|\mathcal{B}|}{(1-\beta)^9\gamma^3\varepsilon^4}\right)$$

episodes of length $\mathcal{O}((1-\beta)^{-1})$. *In the* ν-reset *access model, there is a setting of parameters such that Algorithm 2 when given access to* $\mathcal{W}$ *produces a policy* $\overline{\pi}$ *such that with probability* $1-\delta$, *we have*

$$V^* - V^{\overline{\pi}} \le \frac{D_\infty \mathcal{E}_\nu}{(1-\beta)^2} + \varepsilon,$$

while sampling

$$\mathcal{O}\left(\frac{D_\infty^5 \log|\mathcal{B}|}{(1-\beta)^{15}\gamma^3\varepsilon^5}\right)$$

episodes of length $\mathcal{O}((1-\beta)^{-1})$.

H Proofs for boosting for reinforcement learning

Proof of Theorem 22. The proof here closely follows that of Theorem 7 in [BHS22], and we only indicate the necessary departures. Since we utilize the outer algorithm from previous work, the associated guarantees naturally carry over. The departure comes from our substitution of the *internal boosting* procedure in Algorithm 2 with Algorithm 1; in fact, in its place [BHS22] use a result of [HS21] which can be seen as a generalization of [KK09], e.g., it uses fresh samples for every round of boosting, to other non–binary action sets preserving its $\approx 1/\varepsilon^4$ sample complexity. In this view, the improvement in sample complexity by using the present paper's apprach seems natural.

For the episodic model, applying the second part of Theorem 9 in [BHS22], while noting the smoothness of V^π, and combining the result with Lemma 18 and Lemma 11 in [BHS22], we have with probability $1 - T\delta$

$$V^* - V^{\bar{\pi}} \le \frac{C_\infty \mathcal{E}}{1-\beta} + \frac{4C_\infty^2}{(1-\beta)^3 T} + \frac{C_\infty}{1-\beta}\left(\max_{\pi\in\Pi, t\in[T]} \mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}[\widehat{Q}^\top(\pi(\cdot|s) - \pi'_t(\cdot|s))]\right),$$

where π'_t is the output of the internal boosting algorithm in Line 4 of Algorithm 2, and $\mathcal{D}_t$ is the distribution produced by Algorithm 3 with π_{t-1} selected as the policy for execution. Here, by explicit calculation, noting $a' \sim \text{Unif}(\mathcal{A})$, one may verify for any t that

$$\mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}[y|s] = \frac{1-\beta}{2}\mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}[\widehat{Q}(+1) - \widehat{Q}(-1)|s].$$

Recall that $\mathcal{A} = \{\pm 1\}$, and hence $\pi : \mathcal{S} \to \Delta(\{\pm 1\})$. Hence, for any policy pair π_1, π_2 we have

$$(1-\beta)\mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}[\widehat{Q}^\top(\pi_1(\cdot|s) - \pi_2(\cdot|s))] = \mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}\mathbb{E}_{a_1\sim\pi_1(s)}\mathbb{E}_{a_2\sim\pi_2(s)}[y(a_1 - a_2)].$$

Therefore, all we need to ensure is that output of Algorithm 1 as instantiated in Algorithm 2 every round has an excess correlation gap over the best policy Π no more that $(1-\beta)^2\varepsilon/C_\infty$, which Algorithm 1 assures us can be accomplished with $\mathcal{O}\left(\frac{C_\infty^3 \log|\mathcal{B}|}{(1-\beta)^6\gamma^3\varepsilon^3}\right)$ samples. The total number of samples is $T = \mathcal{O}\left(\frac{C_\infty^2}{(1-\beta)^3\varepsilon}\right)$ times greater.

https://doi.org/10.52202/079017-3190

Similarly, for the ν-reset model, applying the first part of Theorem 9 in [BHS22], and combining the result with Lemma 19 and Lemma 11 in [BHS22], we have

$$V^* - V^{\bar{\pi}} \leq \frac{D_\infty \mathcal{E}_\nu}{(1-\beta)^2} + \frac{2D_\infty}{(1-\beta)^3\sqrt{T}} + \frac{D_\infty}{(1-\beta)^2}\left(\max_{\pi\in\Pi, t\in[T]} \mathbb{E}_{(s,\widehat{Q},y)\sim\mathcal{D}_t}[\widehat{Q}^\top(\pi(\cdot|s) - \pi'_t(\cdot|s))]\right) + \frac{96D_\infty}{(1-\beta)^3\sqrt{P}}\log\frac{1}{\delta}$$

Again, we need to ensure is that output of Algorithm 1 as instantiated in Algorithm 2 every round has an excess correlation gap over the best policy Π no more that $(1-\beta)^3\varepsilon/D_\infty$, which Algorithm 1 assures us can be accomplished with $\mathcal{O}\left(\frac{D_\infty^3 \log|\mathcal{B}|}{(1-\beta)^9\gamma^3\varepsilon^3}\right)$ samples. The total number of samples is $T = \mathcal{O}\left(\frac{D_\infty^2}{(1-\beta)^6\varepsilon^2}\right)$ times greater. □

I Proofs for agnostic learning of halfspaces

Proof of Theorem 8. Using an approximation result of [KOS04], we observe that ERM on the Fourier basis $\chi_S(x) = \prod_{i\in S} x_i$, namely parities on subsets S, can be used to produce a weak learner. This result guarantees that an n-dimensional halfspace can be approximated with uniform weighting on the hypercube to ε^2 ℓ_2-error using degree-limited $\mathcal{B}_{n,d} = \{\pm\chi_S : |S| \leq d\}$ as a basis, where $d = 20\varepsilon^{-4}$. As a result, at least one $h \in \mathcal{B}_{n,d}$ must have high correlation.

Lemma 23 (Lemma 5 in [KMV08])**.** *Let $\mathcal{D}$ be any data distribution over $\{\pm1\}^n \times \{-1,1\}$ with marginal distribution Unif$(\{\pm1\}^n)$ on the features. For any fixed ε and $d = 20\varepsilon^{-4}$, there exists some $h \in \mathcal{B}_{n,d}$ such*

$$\mathrm{corr}_{\mathcal{D}}(h) \geq \frac{\max_{c\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}}(c) - \varepsilon}{n^d}$$

The result follows directly from the preceding lemma, which provides a weak learner for the task, and Theorem 4. We note that $|\mathcal{B}_{n,d}| < n^d$ and $\gamma = n^{-d}$, so

$$\frac{\log|\mathcal{B}_{n,d}|}{\gamma^3\varepsilon^3} \leq \frac{dn^{3d}\log(n)}{\varepsilon^3}.$$

□

J Additional experimental details

For all algorithms, we perform a grid search on the number of boosting rounds with $T \in \{25, 50, 100\}$. For Algorithm 1 we search over $\sigma \in \{0.1, 0.25, 0.5\}$ as well. Rather than using a fixed η, our implementation uses an adaptive step-size scheme proportional to the empirical correlation on the current relabeled distribution. Our experiments were performed using the fractional relabeling scheme stated in [KK09], intended to reduce the stochasticity the algorithm is subject to. In particular, rather than sampling labels, we provide both (x, y) and $(x, -y)$ in our dataset with weights $\frac{1+w^t(x,y)}{2}$ and $\frac{1-w^t(x,y)}{2}$ respectively. For the baselines, $w^t(x,y) = -\phi'_{\text{mad}}(H_t y) = \min(1, \exp(-H_t(x)y))$, where ϕ_{mad} is the Madaboost potential [Dom00]. For the implementation of the proposed algorithm, for greater reproducibility, we use the weighting function $w^t(x,y) = (1-\sigma)\phi'(H_{t-1}(x)y) - \phi'(H_t(x)y)$, which is analytically equivalent to computing the expectation of $p_t(x,y,\eta')$ in Algorithm 1 over $\eta' \sim \text{Unif}[0,\eta]$. The runtime used for all experiments was a Google Cloud Engine VM instance with 2 vCPUs (Intel Xeon 64-bit @ 2.20 GHz) and 12 GB RAM.

K Guide for practical adaptation of Algorithm 1

For many hyperparamters, our theory provides strong clues (the link between η and σ). Briefly, in practice, given a fixed dataset with m_{total} data points, there are two parameters we think a practitioner should concern herself with: the mixing parameter σ and the number of rounds of boosting T, while the rest can be determined, or at least guessed well, given these. The first choice σ dictates the relative

weight ascribed to reused samples across rounds of boosting, while T, apart from determining the running time and generalization error, implicitly limits the algorithm to using m_{total}/T fresh samples each round. For η, one may use the adaptive step-size schedule suggested in the previous section, which also obviates the difficulty of selecting γ. Similarly, our branching criteria, whose necessity is explained in Appendix A, although theoretically both sound and necesssary, is overly conservative. In practice, we believe choosing the better of $-\text{sign}(H_t)$ and W_t whichever produces the greatest empirical distribution on the relabeled distribution will perform best.

https://doi.org/10.52202/079017-3190

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: All claims are backed by theoretical results.

   Guidelines:

   - The answer NA 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 NA 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: All assumptions are stated explicitly. Further, the appendix contains a separate section on limitations of the present work.

   Guidelines:

   - The answer NA 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: Please see the paper.

Guidelines:

- The answer NA 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: [Yes]

   Justification: Please see the additional materials.

   Guidelines:

   - The answer NA 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?

https://doi.org/10.52202/079017-3190

Answer: [Yes]

Justification: Please see the additional materials.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.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://nips.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, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Please see the additional materials and the experiments section for details.

   Guidelines:

   - The answer NA 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: [Yes]

   Justification: Please see the table in the main paper.

   Guidelines:

   - The answer NA 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: [Yes]

   Justification: Please see the appendix.

   Guidelines:

   - The answer NA 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: It does.

   Guidelines:

   - The answer NA 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: [Yes]

    Justification: The present work falls within the realm of foundational research, and is in its basic nature theoretical.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA 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.

https://doi.org/10.52202/079017-3190

- 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., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Not applicable.

    Guidelines:

    - The answer NA 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: [Yes]

    Justification: All datasets used are cited.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: No new assets are introduced in the present work.

    Guidelines:

    - The answer NA 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: [NA]

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

    Guidelines:

    - The answer NA 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: [NA]

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

    Guidelines:

    - The answer NA 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.

 https://doi.org/10.52202/079017-3190