Sample-Optimal Agnostic Boosting with Unlabeled Data

Udaya Ghai [1] **Karan Singh** [2]

Abstract

Boosting provides a practical and provably effective framework for constructing accurate learning algorithms from inaccurate *rules of thumb*. It extends the promise of sample-efficient learning to settings where direct Empirical Risk Minimization (ERM) may not be implementable efficiently. In the realizable setting, boosting is known to offer this computational reprieve without compromising on sample efficiency. However, in the agnostic case, existing boosting algorithms fall short of achieving the optimal sample complexity.

We highlight a previously unexplored avenue of improvement: unlabeled samples. We design a computationally efficient agnostic boosting algorithm that matches the sample complexity of ERM, given polynomially many additional unlabeled samples. In fact, we show that the total number of samples needed, unlabeled and labeled inclusive, is never more than that for the best known agnostic boosting algorithm – so this result is never worse – while only a vanishing fraction of these need to be labeled for the algorithm to succeed. This is particularly fortuitous for learning-theoretic applications of agnostic boosting, which often take place in the *distribution-specific* setting, where unlabeled samples can be availed for free. We also prove that the resultant guarantee is resilient against mismatch between the distributions governing the labeled and unlabeled samples. Finally, we detail an application of this result in reinforcement learning.

1. Introduction

The methodology of boosting, starting with the work of Schapire (1990), has deep roots both in the practice and theory of machine learning. In practice, combining classifiers trained on adaptively weighted data points to highlight past mistakes has proven to be a powerful idea. In theory, boosting provides a highly general and provably efficient framework to convert learning algorithms that are ever so slightly better than random into accurate learners. The historical origins of boosting began with the realization that although Empirical Risk Minimization (ERM), that is, the act of choosing the hypothesis that best fits the observed data, provides a general scheme for statistical learning, for many hypothesis classes it may not be approachable in terms of computational efficiency. Often instead, in theory and practice, it is possible to quickly construct weakly accurate *rules-of-thumb* (or weak learners). Can one use such weak learners to construct an accurate learning algorithm? Boosting provides an affirmative answer to this question by giving a principled way of combining many weak learners, each incrementally focusing on regions of the sample space where others have failed. Part of the original motivation here, as laid out by Kearns & Valiant (1994), was that such an aggregation of hypotheses would no longer be in-class (in modern parlance, *improper*) and hence could potentially circumvent the computational hardness of direct ERM. Interestingly, Kearns & Valiant (1994) were interested in the *representation-independent* hardness of learning, that is, the hardness that persists independently of algorithmic parameterization. However, the main impact of the boosting methodology has been as a *positive* theory, that is, in constructing powerful new learning algorithms, a view that the present work also takes. A natural follow-up question is: *Does the computational reprieve that boosting provides come at a cost?*

Here, we are primarily concerned with sample efficiency. In the *realizable* case, that is, when a perfect classifier exists, the general answer is no. This serves to reaffirm our faith in the realizable boosting framework. The celebrated Adaboost algorithm (Freund & Schapire, 1997), for instance, achieves the same (near-optimal) sample complexity as ERM for VC classes [1]. In fact, Green Larsen & Ritzert (2022) recently showed that for a carefully designed variant the equivalence holds even up to logarithmic terms.

*Equal contribution [1]Amazon, NYC [2]Tepper School of Business, Carnegie Mellon University. Correspondence to: Karan Singh <karansingh@cmu.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

[1]But also see the note in Section 2.1 in Blanc et al. (2024) for potentially a gap given arbitrary weak learners.

Let us now move to the *agnostic* case, where we make no assumptions about the conditional distribution of labels given features; in many ways, this has become the default stance in machine learning. Firstly, the classical notion of realizable weak learning requires that weak learners attain a $0/1$ loss strictly better than half, or equivalently, a correlation strictly better than zero on signed labels. However, such notions are unattainable in the agnostic setting. Early works on agnostic boosting using weak learning definitions that mended this issue did not produce results that were additively competitive with the underlying hypothesis class, thus a direct comparison to ERM remained elusive. Following this, Kanade & Kalai (2009) gave a computationally efficient agnostic boosting that in fact achieves ε-excess population loss compared to the best-in-class hypothesis, mirroring the guarantee ERM provides when coupled with generalization bounds (for example, in Shalev-Shwartz & Ben-David (2014)). However, the sample complexity of their boosting algorithm scales as $(\log|\mathcal{H}|)/\varepsilon^4$, where $\mathcal{H}$ is the (let us say for now, finite[2].) hypothesis class. A recent result of Ghai & Singh (2024) provides a booster that requires $(\log|\mathcal{H}|)/\varepsilon^3$ samples. *However, both exhibit significantly higher sample complexity than the familiar* $(\log|\mathcal{H}|)/\varepsilon^2$ *required for ERM, which is also optimal.*

Although agnostic boosting has seen considerable recent progress both in methodology, for example, through extensions to online (Brukhim et al., 2020) and multiclass (Raman & Tewari, 2022) settings, and in finding new applications (Brukhim et al., 2022; Kothari & Livni, 2018), this limitation persists.

In this work, we highlight an unexplored avenue of improvement: *unlabeled* samples. As usual, to measure the statistical efficiency of the learner, we shall keep track of the number of labeled samples it needs. However, we also imbue the learner with the ability to additionally draw polynomially many unlabeled samples from the underlying feature distribution. Under this stipulation, our main result is a computationally efficient agnostic boosting algorithm with (labeled) sample complexity scaling as $(\log|\mathcal{H}|)/\varepsilon^2$, and therefore for the first time matching that of ERM.

This is fortuitous and relevant both on practical and theoretical fronts. In practice, it is widely acknowledged that, in many domains, unlabeled samples are substantially cheaper to collect than labeled ones. Furthermore, as exemplified by previous works, many applications of agnostic boosting in learning theory take place in *distribution-specific* settings where unlabeled samples can be drawn for free. Here, we expressly utilize the fact that the distribution on features is parsimonious, for example, uniformly distributed or Gaussian, to guarantee that simple classifiers can act as weak learners for the hypothesis class of interest, utilizing tools from L_1-approximation theory. Since in these settings, the feature distributions are explicitly noted, even closing the aforementioned gap will obtain results that are no better than what we get. We offer further refinements and applications of our main result, as we detail next.

[2]Purely for ease of presentation, in the introduction, we stick to finite classes and drop dependencies in the weak learning edge and the failure probability, later denoted by γ and δ respectivey. In fact, our results hold for infinite VC classes. See Theorem 3.1

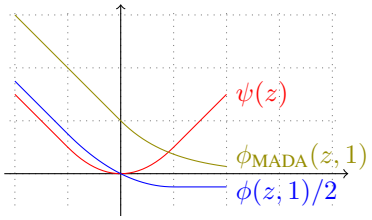

Figure 1. The potential $\phi(z,1)/2$ compared against the Madaboost potential $\phi_{\text{MADA}}(z,1)$ (Domingo, 2000), along with the the Huber loss $\psi(z)$.

1.1. Overview of contributions and techniques

Sample-optimal boosting with unlabeled data. Our first and main result is a computationally efficient agnostic boosting algorithm that produces a classifier with ε excess error, given $(\log|\mathcal{H}|)/\varepsilon^2$ labeled and $(\log|\mathcal{H}|)/\varepsilon^4$ unlabeled samples. Our labeled sample complexity is essentially optimal.

Using unlabeled data to accelerate boosting is a novel idea. Our work also differs qualitatively, in that it does not require additional assumptions, compared to semi-supervised learning methods (Chapelle et al., 2006), which often require a tight clustering of data or smoothness of labels.

Despite differing in algorithmic techniques, previous works (Kanade & Kalai, 2009; Ghai & Singh, 2024) on agnostic boosting almost invariably use the Madaboost potential (Domingo, 2000) or a closely related variant:

$$\phi_{\text{MADA}}(z,y)=\begin{cases}e^{-zy} & \text{if } zy\geq 0,\\ 1-zy & \text{if } zy<0,\end{cases}$$

where z is a real-valued prediction of the discrete label y.

Our key innovation is to design a new *bivariate* potential

$$\phi(z,y)=-yz+\begin{cases}|z|-\frac{1}{2} & \text{if } |z|>1,\\ \frac{1}{2}z^2 & \text{if } |z|\leq 1.\end{cases},$$

whose derivative (with respect to z) is decomposable as one term exclusively dependent on the label, and the other exclusive involving the features. Thus, by linearity of expectation, the population analogues of these parts can be estimated separately with labeled and unlabeled data. Furthermore, it is essential that the labeled data part, or formally its derivative, does not depend on the ensemble whose value is being

assessed; this allows us to reuse the same set of labeled data across all rounds of boosting unlike Kanade & Kalai (2009), without needing a uniform convergence argument as in Brukhim et al. (2020), or a martingale argument as in Ghai & Singh (2024).

This non-conforming potential function violates some key tenets of prior work. For example, the centerpiece of the Madaboost potential is that it, or formally, its functional derivative, does not downweight points that are misclassified by the ensemble. This property is best captured in the definition of a conservative relabeling in Kanade & Kalai (2009). Intuitively, it makes sense; one wants not to withdraw any focus from wrongly classified data points. Not only is this false for us, it is incompatible with the requirement of having a (smooth) separable potential function of the manner we have just described. This can be seen in Figure 1: the negative slope of the potential, always between 0 and 1, specifies the weight of each sample. The derivative of Madaboost for the negative domain is always -1, but for us this is not true, and our potential curves upward between -1 and 0.

Improved unlabeled sample efficiency. Since our main innovation, as described above, is orthogonal to previous approaches, it can be layered on top of existing algorithmic techniques.

Specifically, applying techniques from Ghai & Singh (2024) selectively to the part of potential that arises from unlabeled data, we further reduce the number of unlabeled samples needed to $\log(|\mathcal{H}|)/\varepsilon^3$. With this result, our total number of samples (labeled and unlabeled) matches the number of labeled samples needed by the previous best result, yet for us, only a vanishing, concretely ε, fraction of these needs to be labeled.

Resilience to covariate shift. Since, in addition to the labeled dataset, our algorithm has access to a stream of unlabeled examples, a natural practical concern may be that the underlying law governing the latter's generation may not exactly match the distribution of features in labeled examples. Unlike realizable learning where every sample helps to narrow down to the correct classifier, agnostic learning is often brittle against changes in the covariate distribution, since no classifier is the best in all regions of the feature space. However, we show that if there is a mismatch between the labeled and unlabeled distributions available to learner, our learner still succeeds in learning an *arbitrarily accurate* classifier as long as these distributions have the same support on the feature space. Moreover, the labeled sample complexity is unaffected by such a mismatch.

We point out that for the covariate shift setting, one must suitably generalize the measure of the progress being made in each round of boosting. This requires a few changes, the first among which is that the potential function is no longer the population (expectation) version of a scalar potential. We keep track of the progress on the labeled and unlabeled distributions separately.

Applications. We apply our results to boosting for reinforcement learning, where we reduce the required number of reward-annotated episodes, and learning halfspaces.

1.2. Related work

The early theory of boosting was developed in a sequence of papers (Freund, 1995; Schapire & Singer, 1998; Bartlett et al., 1998) starting with Schapire (1990) leading to a breakthrough in the form of Adaboost (Freund & Schapire, 1997). Even earlier the possibility of boosting was posed in Kearns (1988); Kearns & Valiant (1994). A comprehensive survey can be found in Schapire & Freund (2013). Early boosting algoritms were quite sensitive to noise (Long & Servedio, 2008; Domingo, 2000). To mitigate this, boosting was then studied in the random classification noise model (Diakonikolas et al., 2021; Kalai & Servedio, 2003). Agnostic boosting started with Ben-David et al. (2001); Mansour & McAllester (2002); Kalai et al. (2008); Kale (2007); Chen et al. (2016) where new types of weak learners suitable for the agnostic setting were defined. Kanade & Kalai (2009) (see also Feldman (2010)) gave weak learning definition that led to additive excess error guarantees. Boosting in online setting is also well studied (Beygelzimer et al., 2015; Chen et al., 2012; Jung et al., 2017; Brukhim et al., 2020; Raman & Tewari, 2022; Hazan & Singh, 2021). See Alon et al. (2021); Green Larsen & Ritzert (2022); Lyu et al. (2024) for more recent work.

An ostensible alternative to our approach, especially given that realizable boosting achieves a near-optimal sample complexity, is to reduce agnostic boosting to the realizable case à la Hopkins et al. (2022). The concurrent work of da Cunha et al. (2025) pursues this and, in fact, obtains much more refined fat-shattering bounds. Although statistically optimal, this reduction requires pruning the hypothesis class by enumeration while considering all possible labelings of samples, and hence takes exponential time, rivaling ERM. In contrast, the present work offers computationally efficient algorithms that run in polynomial time, a requirement that has been central to the theory of boosting since its origin in computational learning theory.

2. Problem setting

We consider the fundamental setting of binary classification where $\mathcal{X}$ represents the set of feature descriptions, and the possible labels are $\{\pm 1\}$. There is an underlying joint distribution $\mathcal{D}$ over $\mathcal{X} \times \{\pm 1\}$, which, while unknown to the learner, is crucial to determining its performance. Let $\mathcal{D}_{\mathcal{X}}$

be the marginal distribution $\mathcal{D}$ induces on the feature space $\mathcal{X}$. Given any binary classifier $h : \mathcal{X} \to \{\pm 1\}$, its success on $\mathcal{D}$ can be measured as

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

Correlation can readily be translated to the more commonly used metric, the $0/1$-loss $l^{0/1}_{\mathcal{D}}(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}} \left[\mathbf{1}_{y\neq h(x)}\right]$, as $\mathrm{corr}_{\mathcal{D}}(h) = 1 - 2l^{0/1}_{\mathcal{D}}(h)$.

Consider a hypothesis class $\mathcal{H} \subseteq \mathcal{X} \to \{\pm 1\}$ against which the learner aims to be competitive. The objective of the learner is to produce a hypothesis $\overline{h} : \mathcal{X} \to \{\pm 1\}$ such that with probability at least $1 - \delta$, $\mathrm{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}}(h) - \varepsilon$, where ε, δ are pre-specified error tolerances. A crucial remark here is that this final hypothesis may not belong to the hypotheses class $\mathcal{H}$. The learners who are given such flexibility are called *improper*, and all known boosting algorithms fall into this class.

Definition 2.1 (Agnostic Weak Learner)**.** For any $\varepsilon_0, \delta_0 > 0$, a γ-agnostic weak learner for a hypothesis class $\mathcal{H}$ and a base class $\mathcal{B}$ draws $m(\varepsilon_0, \delta_0)$ independently and identically distributed samples from any distribution $\mathcal{D}'$ supported on $\mathcal{X} \times \{\pm 1\}$ and outputs a base hypothesis $\mathcal{W} \in \mathcal{B}$ such that with probability at least $1 - \delta_0$,

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

This definition of agnostic weak learning was introduced in Kanade & Kalai (2009), where it is noted that typically $m(\varepsilon, \delta) = O(\log(|\mathcal{B}|/\delta)/\varepsilon^2)$. Mirroring the presentation of results in Kanade & Kalai (2009); Brukhim et al. (2020); Ghai & Singh (2024), we present the formal statement of results for fixed ε_0, δ_0. In this way, the magnitudes of contribution to the final error for the boosting algorithm and the weak learner are made distinct.

Finally, a further refinement of the agnostic boosting framework is the *distribution-specific* setting (Kanade & Kalai, 2009; Feldman, 2010), in which the set of input distributions to the weak learner are constrained so that their marginals on the feature space match that of the true underlying distribution. Thus, any regularity present in the feature distribution $\mathcal{D}_{\mathcal{X}}$, e.g., if it follows a uniform or a Gaussian distribution, is also made available to the weak learner, which makes the design of such weak learners easier. Our results will also apply under this restriction. Like in previous work, our main algorithm works by *relabeling examples*; there is no need to adaptively reweigh them. This ensures that, for any sample fed to the weak learner, the overall marginal distribution follows $\mathcal{D}_{\mathcal{X}}$.

3. The algorithm and the main result

Our main result and its proof are completely contained in this section. We describe some notation and essential algorithmic ingredients in the following.

Notation. Let $h^* = \arg\max_{h\in\mathcal{H}} \mathrm{corr}_{\mathcal{D}}(h)$ be the best-in-class hypothesis. For brevity of notation, for any finite set $D \subseteq \mathcal{X} \times \{\pm 1\}$, we denote the empirical average over it by $\widehat{\mathbb{E}}_D[\cdot] = \frac{1}{|D|}\sum_{(x,y)\in D}(\cdot)$. We define $\mathrm{sign}(z)$ as 1 if $z \geq 0$ and -1 otherwise. For a real-valued function f, we take $\mathrm{sign}\, f$ to mean its precomposition with sign.

Potential function. Consider the potential function

$$\phi(z, y) = \psi(z) - yz \,, \tag{1}$$

where $\psi(z)$ is the Huber loss (Huber, 1992):

$$\psi(z) = \begin{cases} |z| - \frac{1}{2} & \text{if } |z| > 1, \\ \frac{1}{2}z^2 & \text{if } |z| \leq 1. \end{cases}$$

Since we never differentiate $\phi(z, y)$ with respect to y, let $\phi'(z, y) = \frac{\partial\phi(z,y)}{\partial z}$ and $\phi''(z, y) = \frac{\partial^2\phi(z,y)}{\partial^2 z}$. To measure the progress of any *real-valued* $H : \mathcal{X} \to \mathbb{R}$, consider the population potential

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

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

The quantity $\Phi'_{\mathcal{D}}(H, h)$ is the directional derivative of $\Phi_{\mathcal{D}}(H)$ on h.

Description of the Algorithm. Algorithm 1 roughly follows the potential based boosting framework of Kanade & Kalai (2009). The algorithm simulates the process where the true label is kept with probability $-\phi'(H_t(x), y)/2$ and chosen randomly otherwise. As can be seen in Figure 1, the probability of flipping increases monotonically in the magnitude of $yH_t(x)$, so the more certain H_t is of the correct label, the closer to uniformly random the label will be in $\mathcal{D}_t$ for the weak learner. Since predicting on random labels is impossible, this randomized relabeling increases the relative importance of data that H_t does not predict accurately.

The main trick we employ in this work is that the potential $\phi(z, y)$ in (1) is split into two parts such that the first part $\psi(z)$ has *no dependence on the label*, and hence can be estimated via unlabeled examples. The second part $-yz$ is *linear* in z, hence the derivative has no dependence on z. As a result estimating this does not depend on H_t, but just a weak hypothesis. Since this is a simple class, the samples required for this estimation are small and we can

use uniform convergence to assure concentration across all boosting rounds. The formal concentration argument can be seen in Lemma 3.3. One interesting observation, is that in this construction, samples from the labeled part of the distribution are never relabeled. Line 6 of Algorithm 1 can be interpreted as providing a regularization, wherein predictions on the unlabeled data are pushed towards 0 because $p_t(x) \leq \frac{1}{2}$ if and only if $H_t(x) \geq 0$.

The relabeling is designed so correlation on the relabeled distribution corresponds to the derivative of a population potential (see Lemma 3.3). As such, a weak learner produces a hypothesis that has nonnegligeable correlation with the (negative) functional gradient of the population potential $\Phi_{\mathcal{D}}(H_t)$. With properly chosen η, this assures a descent in potential as long as the weak learner has sufficient edge on the current distribution (Case A of Theorem 3.1 and line 10 of Algorithm 1). However, this need not be the case because we have access to an *agnostic* weak learner, and it's possible that no $h \in \mathcal{B}$ performs well on $\mathcal{D}_t$. If this is the case and if $\Phi'_{\mathcal{D}}(H_t, \operatorname{sign}(H_t)) \geq \varepsilon$ we are in Case B of Theorem 3.1. In this case, this condition on the derivative assures us that $h_t = -\operatorname{sign} H_t$ is also a descent direction (line 12 of Algorithm 1). Now, because the potential is bounded from below, only a certain number of such descent steps can occur. Eventually, we must reach a point where neither Case A nor Case B holds. Here $\Phi'_{\mathcal{D}}(H_t, \operatorname{sign}(H_t))$ is small and there is no $h \in \mathcal{B}$ that performs well on $\mathcal{D}_t$. Stitching these together with Lemma 3.5 relating the correlations to the potential gradient can be used to provide the result.

3.1. Main result on sample-optimal agnostic boosting

Theorem 3.1 (Main theorem)**.** *For any $\varepsilon, \delta > 0$, there is an instantiation of parameters such that $\eta = \mathcal{O}(\gamma^2\varepsilon)$, $T = \mathcal{O}(1/\gamma^2\varepsilon^2)$, $\tau = \mathcal{O}(\gamma\varepsilon)$, $S = \mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^2\varepsilon^2)$, $U = \mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^2\varepsilon^2)$, $S_0 = \mathcal{O}(1/\varepsilon^2)$, $m = m(\varepsilon_0, \delta_0) + \mathcal{O}(1/\gamma^2\varepsilon^2)$ for which Algorithm 1 guarantees with probability $1 - \delta - T\delta_0$ that*

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

During its execution, Algorithm 1 makes $T = \mathcal{O}(1/\gamma^2\varepsilon^2)$ calls to the weak learner, and samples $S + S_0 = \mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^2\varepsilon^2)$ labeled samples and $TU = \mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^4\varepsilon^4)$ unlabeled samples.

3.2. Proof of the main result

To maintain the continuity of presentation, our organization and notation closely mirror Kanade & Kalai (2009). We will use the following properties of ϕ and ψ.

Proposition 3.2. *$|\psi'(z)| \leq 1$ and $z\psi'(z) \geq 0$ for all $z \in \mathbb{R}$. For all $y \in \{\pm 1\}$, $\phi(\cdot, y)$ is continuously differentiable, 1-smooth, and satisfies $\phi(0, y) - \min_z \phi(z, y) \leq 1/2$.*

Algorithm 1 Agnostic Boosting with Unlabeled Data

1: **Inputs:** Samplers for labeled data from $\mathcal{D}$ and unlabeled data from $\mathcal{D}_{\mathcal{X}}$, γ-agnostic weak learning oracle $\mathcal{W}$, parameters $\eta, T, \tau, S, U, S_0, m$.
2: Initialize a zero hypothesis $H_1 = \mathbf{0}$.
3: Sample S *labeled* examples to create dataset $\widehat{D}$.
4: **for** $t = 1$ to T **do**
5: Sample U *unlabeled* examples to create dataset $\widehat{D}_t$.
6: Construct a resampling distribution $\mathcal{D}_t$ that chooses between steps A and B with equal probability.
A. Return (x, y) picked uniformly from $\widehat{D}$.
B. Pick x uniformly from $\widehat{D}_t$, and return $(x, \widehat{y})$, where the pseudo-label $\widehat{y}$ is chosen as
$$\widehat{y} = \begin{cases} +1 & \text{with probability } p_t(x) = \frac{1 - \psi'(H_t(x))}{2}, \\ -1 & \text{with remaining probability.} \end{cases}$$
7: Sample m times from $\mathcal{D}_t$ to create dataset $\widehat{D}'_t$.
8: Call the weak learner on $\widehat{D}'_t$ to get $W_t = \mathcal{W}(\widehat{D}'_t)$.
9: **if** $\operatorname{corr}_{\widehat{D}'_t}(W_t) = \sum_{(x,\widehat{y}) \in \widehat{D}'_t} \widehat{y} W_t(x) > \tau$ **then**
10: Update $H_{t+1} = H_t + \eta W_t / \gamma$.
11: **else**
12: Update $H_{t+1} = H_t - \eta \operatorname{sign}(H_t)_t$.
13: **end if**
14: **end for**
15: Sample S_0 *labeled* examples to create dataset $\widehat{D}_0$.
16: **Output** $\overline{h} = \underset{h \in \{\operatorname{sign}(H_t) : t \in [T]\}}{\arg\max} \sum_{(x,y) \in \widehat{D}_0} y h(x)$.

Proof of Proposition 3.2. Let us begin by noting that $\psi'(z) = \operatorname{sign}(z) \min\{1, |z|\}$ from which the first two properties can be seen. From this, all but the last properties follow. The last part can be verified by elementary calculations. □

We show that $\Phi'_{\mathcal{D}}(H, h)$ can be estimated efficiently on the base class $\mathcal{B}$.

Lemma 3.3. *There exists a universal constant $C > 0$ such that with probability $1 - \delta$, for all $t \in [T], h \in \mathcal{B} \cup \{h^*\}$,*

$$\left| \frac{1}{2} \Phi'_{\mathcal{D}}(H_t, h) + \operatorname{corr}_{\mathcal{D}_t}(h) \right| \leq \varepsilon_{Gen} := C \sqrt{\frac{\mathrm{VC}(\mathcal{B}) + \log \frac{1}{\delta}}{\min\{S, U\}}}.$$

Proof of Lemma 3.3. By the definition of $\mathcal{D}_t$, we have that

$$\operatorname{corr}_{\mathcal{D}_t}(h) = \frac{1}{2} \widehat{\mathbb{E}}_{\widehat{D}}[y h(x)] - \frac{1}{2} \widehat{\mathbb{E}}_{\widehat{D}_t}[\psi'(H_t(x)) h(x)],$$

where we use the fact that Line 6.B in Algorithm 1 ensures $\mathbb{E}[\widehat{y}|x] = -\psi'(H_t(x))$. Since $\widehat{D}$ and $\widehat{D}_t$ are composed of IID draws from $\mathcal{D}$ and $\mathcal{D}_{\mathcal{X}}$ respectively, we use the following uniform convergence result (e.g., see Anthony & Bartlett (2009)), originally due to Talagrand (1994).

Theorem 3.4 ((Talagrand, 1994)). *Fix a hypothesis class $\mathcal{B} \subseteq \mathcal{X} \to \{\pm 1\}$, and distribution $\mathcal{D}$ over $\mathcal{X} \times \{\pm 1\}$. There is a universal constant $C \geq 0$ such that with probability $1-\delta$, for all $h \in \mathcal{B}$, it holds*

$$\left| \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yh(x)] - \frac{1}{m}\sum_{i=1}^{m} y_i h(x_i) \right| \leq C\sqrt{\frac{\mathrm{VC}(\mathcal{B}) + \log\frac{1}{\delta}}{m}},$$

where $\{(x_i, y_i)\}_{i\in[m]}$ are sampled IID from $\mathcal{D}$.

Hence, for some constant $C \geq 0$ we have with probability $1-\delta$ that $|\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \mathrm{corr}_{\mathcal{D}}(h)|$ and $|\widehat{\mathbb{E}}_{\widehat{D}_t}[\psi'(H_t(x))h(x)] - \mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}}[\psi'(H_t(x))h(x)]|$ are at most ε_{Gen}. Since $\phi'(z,y) = \phi'(z) - y$, and hence

$$\Phi'_{\mathcal{D}}(H_t, h) = \mathop{\mathbb{E}}_{x\sim\mathcal{D}_{\mathcal{X}}}[\psi'(H_t(x))h(x)] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yh(x)],$$

completes the proof. Including h^* changes the VC dimension by a constant (Eisenstat & Angluin, 2007). □

Finally, we will use the following lemma that upper bounds the correlation gap, which is our ultimate concern, by the difference in the directional derivative of the potential.

Lemma 3.5. *For any real-valued classifier $H : \mathcal{X} \to \mathbb{R}$, we have* $\mathrm{corr}_{\mathcal{D}}(h^*) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}(H)) \leq \Phi'_{\mathcal{D}}(H, \mathrm{sign}(H)) - \Phi'_{\mathcal{D}}(H, h^*)$.

Proof. We start by noting that

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

where in the last line we use the fact that $y^2 = 1$.

Consider any (x,y) such that $yH(x) > 0$: Here $y(h^*(x) - \mathrm{sign}(H(x))) < 0$. Furthermore, since y and $H(x)$ have the same sign, so do y and $\psi'(H(x))$ by Proposition 3.2, and hence $(1 - y\psi'(H(x))) \leq 1$. Similarly, whenever $yH(x) < 0$: Then $y(h^*(x) - \mathrm{sign}(H(x))) > 0$, and y and $\psi'(H(x))$ have opposite signs that imply $(1 - y\psi'(H(x))) \geq 1$.

Now the claim follows as

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

This wraps up the lemma. □

We are now ready to prove the main result.

Proof of Theorem 3.1. Let us dispense with the random events at once. The success of Lemma 3.3, the event that $\max_{t\in[T]} |\mathrm{corr}_{\widehat{D}_0}(\mathrm{sign}\, H_t) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}\, H_t)| \leq \varepsilon/10$, and $\max_{t\in[T]} |\mathrm{corr}_{\widehat{D}'_t}(W_t) - \mathrm{corr}_{\mathcal{D}_t}(W_t)| \leq \gamma\varepsilon/10$ can be ensured with probability $1-\delta$ by a simple application of Hoeffing's inequality and union bound given the setting of m and S_0 in the statement of the theorem. Similarly, $\varepsilon_{\text{Gen}} \leq \gamma\varepsilon/10$ holds in Lemma 3.3 for $S, U = \Omega((\mathrm{VC}(\mathcal{B}) + \log\delta^{-1})/\gamma^2\varepsilon^2)$.

Let $h_t = (H_{t+1} - H_t)/\eta$. Since ϕ is 1-smooth by Proposition 3.2:

$$\begin{aligned}
&\Phi_{\mathcal{D}}(H_{t+1}) - \Phi_{\mathcal{D}}(H_t) \qquad (2)\\
&\leq \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\eta\phi'(H_t(x), y)h_t(x) + \frac{\eta^2(h_t(x))^2}{2}\right] \\
&\leq \eta\Phi'_{\mathcal{D}}(H_t, h_t) + \frac{\eta^2}{2\gamma^2}. \qquad (3)
\end{aligned}$$

Case A: Consider any step t where $\mathrm{corr}_{\widehat{D}'_t}(W_t) > \tau$. Here $h_t = W_t/\gamma$. It follows from Lemma 3.3 that

$$\begin{aligned}
\Phi'_{\mathcal{D}}(H_t, h_t) &\leq -\frac{2\,\mathrm{corr}_{\mathcal{D}_t}(h_t)}{\gamma} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} \\
&\leq -\frac{2\,\mathrm{corr}_{\widehat{D}'_t}(h_t)}{\gamma} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} + \frac{\varepsilon}{5} \\
&\leq -\varepsilon
\end{aligned}$$

where we set $\tau = \gamma\varepsilon$ and $\eta = \gamma^2\varepsilon$. By Equation (2), the potential drops as $\Phi_{\mathcal{D}}(H_{t+1}) - \Phi_{\mathcal{D}}(H_t) \leq -\gamma^2\varepsilon^2/2$.

Case B: Consider any step t where $\mathrm{corr}_{\widehat{D}'_t}(W_t) \leq \tau$ and *crucially* $\Phi'_{\mathcal{D}}(H_t, \mathrm{sign}\, H_t) \geq \varepsilon$. Here $h_t = -\mathrm{sign}\, H_t$. Since $\Phi'_{\mathcal{D}}(H_t, h_t) = -\Phi'_{\mathcal{D}}(H_t, \mathrm{sign}\, H_t) \leq -\varepsilon$, by Equation (2), we have $\Phi_{\mathcal{D}}(H_{t+1}) - \Phi_{\mathcal{D}}(H_t) \leq -\gamma^2\varepsilon^2/2$.

By Proposition 3.2, at initialization, $\Phi_D(\mathbf{0})$ is at most a half away from the minimum. Thus, setting $T = 2/\gamma^2\varepsilon^2$, there must arise an iterate such that neither Case A nor Case B hold. That is, there is some $s \in [T]$ such that $\mathrm{corr}_{\widehat{D}'_s}(W_s) \leq \tau$ and $\Phi'_{\mathcal{D}}(H_s, \mathrm{sign}\, H_s) \leq \varepsilon$. Now using Lemma 3.3 and that the weak learner γ-approximately maximizes correlation (Definition 2.1), we have

$$\begin{aligned}
-\Phi'_{\mathcal{D}}(H_s, h^*) &\leq 2\,\mathrm{corr}_{\mathcal{D}_s}(h^*) + 2\varepsilon_{\text{Gen}} \\
&\leq \frac{2\,\mathrm{corr}_{\mathcal{D}_s}(W_s)}{\gamma} + \frac{2\varepsilon_0}{\gamma} + 2\varepsilon_{\text{Gen}} \\
&\leq \frac{2\,\mathrm{corr}_{\widehat{D}'_s}(W_s)}{\gamma} + \frac{2\varepsilon_0}{\gamma} + \frac{2\varepsilon}{5} \\
&\leq \frac{2\varepsilon_0}{\gamma} + \frac{12\varepsilon}{5}
\end{aligned}$$

where in the last line we recall $\tau = \varepsilon\gamma$ and $\varepsilon_{\text{Gen}} = \varepsilon\gamma/10$.

By Lemma 3.5, we have

$$\begin{aligned}&\operatorname{corr}_{\mathcal{D}}(h^*) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign}(H_s))\\ &\le \Phi'_{\mathcal{D}}(H_s, \operatorname{sign}(H_s)) - \Phi'_{\mathcal{D}}(H_s, h^*)\\ &\le \frac{2\varepsilon_0}{\gamma} + \frac{17\varepsilon}{5}.\end{aligned}$$

To complete the proof, we observe that

$$\begin{aligned}\operatorname{corr}_{\mathcal{D}}(\overline{h}) &\le \operatorname{corr}_{\widehat{D}_0}(\overline{h}) + \varepsilon/10\\ &\le \operatorname{corr}_{\widehat{D}_0}(\operatorname{sign} H_s) + \varepsilon/10\\ &\le \operatorname{corr}_{\mathcal{D}}(\operatorname{sign} H_s) + \varepsilon/5\\ &\le 2\varepsilon_0/\gamma + 18\varepsilon/5,\end{aligned}$$

where we use the fact that $\overline{h}$ maximizes the empirical correlation on the dataset $\widehat{D}_0$. Substituting ε appropriately yields the claim. □

4. Improving unlabeled sample efficiency

In this section, we reduce the number of unlabeled samples needed to $1/\gamma^3\varepsilon^3$, using the data reuse scheme from Ghai & Singh (2024), which is crucially only applied to unlabeled data. The key idea behind the scheme is that since H_t changes by a small amount each time, the change it induces on any twice-continuously differentiable potential is also small. Therefore, the desired relabeling distributions are not too different across rounds, and one may be able to reuse the distribution of past rounds to some extent. This difference is reflected in Algorithm 2 in Line 6 which allows recursive use of unlabeled data from past rounds. To do this, we first construct a twice-continuously differentiable potential using a Pseudo-Huber loss.

$$\phi(z, y) = \psi(z) - yz, \text{ and } \psi(z) = \sqrt{1 + x^2} - 1. \quad (4)$$

Theorem 4.1 (Main theorem with unlabeled data reuse)**.** *For any $\varepsilon, \delta > 0$, there is an instantiation of parameters such that $\eta = \mathcal{O}(\gamma^2\varepsilon/\log|\mathcal{B}|)$, $T = \mathcal{O}(\log|\mathcal{B}|/\gamma^2\varepsilon^2)$, $\tau = \mathcal{O}(\gamma\varepsilon)$, $S = \mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^2\varepsilon^2)$, $U = \mathcal{O}(1/\gamma\varepsilon)$, $S_0 = \mathcal{O}(1/\varepsilon^2)$, $m = m(\varepsilon_0, \delta_0) + \mathcal{O}(1/\gamma^2\varepsilon^2)$ for which Algorithm 2 guarantees with probability $1 - \delta - T\delta_0$ that*

$$\operatorname{corr}_{\mathcal{D}}(\overline{h}) \ge \max_{h\in\mathcal{H}} \operatorname{corr}_{\mathcal{D}}(h) - \frac{3\varepsilon_0}{\gamma} - \varepsilon.$$

During its execution, Algorithm 2 makes $T = \mathcal{O}(\log|\mathcal{B}|/\gamma^2\varepsilon^2)$ calls to the weak learner, and needs $S + S_0 = \mathcal{O}(\log|\mathcal{B}|/\gamma^2\varepsilon^2)$ labeled samples and $TU = \mathcal{O}(\log|\mathcal{B}|/\gamma^3\varepsilon^3)$ unlabeled samples.

Although the above result is always better in terms of the demand for unlabeled samples, it comes at the cost of increased oracle complexity, that is, the number of calls to the weak learner, which now has a $\log|\mathcal{B}|$ factor unlike Theorem 3.1. Again mirroring techniques in Ghai & Singh (2024) provides some mitigation. In particular, in Appendix A.2, we provide a different guarantee for the same algorithm that makes $\mathcal{O}(1/\gamma^2\varepsilon^2)$ calls to the weak learner, while needing $\mathcal{O}(\log|\mathcal{B}|/\gamma^2\varepsilon^2)$ labeled and $\mathcal{O}(\log|\mathcal{B}|/\gamma^3\varepsilon^3 + (\log|\mathcal{B}|)^3/\gamma^2\varepsilon^2)$ unlabeled samples.

Algorithm 2 Agnostic Boosting with Selcetive Reuse of Unlabeled Data

1: **Inputs:** Samplers for labeled data from $\mathcal{D}$ and unlabeled data from $\mathcal{D}_{\mathcal{X}}$, γ-agnostic weak learning oracle $\mathcal{W}$, parameters $\eta, T, \tau, S, U, S_0, m, \sigma$.
2: Initialize a zero hypothesis $H_1 = \mathbf{0}$.
3: Sample S *labeled* examples to create dataset $\widehat{D}$.
4: **for** $t = 1$ to T **do**
5: Sample U *unlabeled* examples to create dataset $\widehat{D}_t$.
6: Construct a resampling distribution $\mathcal{D}'_t$ that picks x uniformly from $\widehat{D}_t$, picks $\widehat{y} \in \{\pm 1\}$ uniformly and returns $(x, \widehat{y})$ if $t = 1$; for $t > 1$, do:
A. With probability $1 - \sigma$, return a sample $(x, \widehat{y})$ from $\mathcal{D}'_{t-1}$.
B. Else return $(x, \widehat{y})$ where x is uniformly chosen from $\widehat{D}_t$, $\eta' \sim \mathrm{Unif}[0, \eta]$, and $\widehat{y}$ is created as

$$\widehat{y} = \begin{cases} +1 & \text{with probability } p_t(x, \eta'),\\ -1 & \text{with remaining probability, where}\end{cases}$$

$$\begin{aligned}p_t(x, \eta') = \frac{1}{2} &- \frac{\sigma\psi'(H_{t-1}(x))}{2(\eta + \sigma)}\\ &- \frac{\eta\psi''(H_{t-1}(x) + \eta' h_{t-1}(x))h_{t-1}(x)}{2(\eta + \sigma)}.\end{aligned}$$

7: Construct a resampling distribution $\mathcal{D}_t$ that chooses between steps A and B with equal probability.
A. Return (x, y) picked uniformly from $\widehat{D}$.
B. Return $(x, \widehat{y})$ sampled uniformly from $\mathcal{D}'_t$.
8: Sample m times from $\mathcal{D}_t$ to create dataset $\widehat{D}'_t$.
9: Call the weak learner on $\widehat{D}'_t$ to get $W_t = \mathcal{W}(\widehat{D}'_t)$.
10: **if** $\operatorname{corr}_{\widehat{D}'_t}(W_t) = \sum_{(x,\widehat{y})\in\widehat{D}'_t} \widehat{y} W_t(x) > \tau$ **then**
11: Update $H_{t+1} = H_t + \eta W_t/\gamma$.
12: **else**
13: Update $H_{t+1} = H_t - \eta \operatorname{sign}(H_t)_t$.
14: **end if**
15: **end for**
16: Sample S_0 *labeled* examples to create dataset $\widehat{D}_0$.
17: **Output** $\overline{h} = \underset{h\in\{\operatorname{sign}(H_t):t\in[T]\}}{\arg\max} \sum_{(x,y)\in\widehat{D}_0} yh(x)$.

5. Resiliency against covariate shift

We consider the setting when the learner has access to a distribution $\mathcal{D}$ supported over $\mathcal{X} \times \{\pm 1\}$ and a different, possibly unrelated, distribution $\mathcal{Q}$ supported over features $\mathcal{X}$. We show that under mild conditions the learner can still produce an arbitrarily accurate classifier. To measure how different $\mathcal{Q}$ and $\mathcal{D}_{\mathcal{X}}$ are, we define $C_{\mathcal{X}} = \|d\mathcal{D}_{\mathcal{X}}/d\mathcal{Q}\|_{\infty}$, which is a uniform upper bound on the Radon-Nikodym derivative of $\mathcal{D}_{\mathcal{X}}$ with respect to $\mathcal{Q}$.

Theorem 5.1 (Main theorem for covariate shift)**.** *For any $\varepsilon, \delta > 0$, there is an algorithm that makes $\mathcal{O}(\mathcal{C}_{\mathcal{X}}/\gamma^2\varepsilon^2)$ calls to the weak learner, samples $\mathcal{O}(\mathrm{VC}(\mathcal{B})/\gamma^2\varepsilon^2)$ labeled samples from $\mathcal{D}$ and $TU = \mathcal{O}((\mathcal{C}_{\mathcal{X}})^3\,\mathrm{VC}(\mathcal{B})/\gamma^4\varepsilon^4)$ unlabeled samples from $\mathcal{Q}$, and outputs $\overline{h}$ such that with probability $1 - \delta - T\delta_0$ that*

$$\mathrm{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h \in \mathcal{H}} \mathrm{corr}_{\mathcal{D}}(h) - \frac{(1 + \mathcal{C}_{\mathcal{X}})\varepsilon_0}{\gamma} - \varepsilon.$$

Recall that ε_0 can be made arbitrarily small by feeding more samples to the weak learner from the empirical distribution. Although this comes at a (polynomial) computational cost, in particular, the weak learner now needs to be more accurate, the sample complexity remains unaffected.

The key step is in Lemma B.3; it proves an analogue of Lemma 3.5 implying that oversampling the part of the potential connected to the Huber loss does not hurt the correlation gap. From there, we set up a non-scalar potential measure to keep track of the progress of the learner.

6. Applications

6.1. Agnostic learning of halfspaces

We illustrate how our method, when used as a black box, agnostically learns halfspaces over the n-dimensional Boolean hypercube under uniform marginals on the features. Since unlabeled samples can be drawn from the uniform distribution at essentially no statistical cost, the added complexity of acquiring unlabeled data becomes solely a computational concern.

This procedure improves upon existing *boosting*-based approaches. In particular, building on Kanade & Kalai (2009), we rely on empirical risk minimization (ERM) over all parities of degree at most $d \approx 1/\varepsilon^4$. Such an ERM rule achieves a weak learner advantage $\gamma = n^{-d}$. As shown in Theorem 6.1 below (proved in Appendix D), this instantiation of our boosting framework reduces the labeled sample complexity from $\mathcal{O}(\varepsilon^{-7}n^{60\varepsilon^{-4}})$ to $\mathcal{O}(\varepsilon^{-6}n^{40\varepsilon^{-4}})$.

Theorem 6.1. *Let $\mathcal{D}$ be any distribution over $\{\pm 1\}^n \times \{\pm 1\}$ with uniform feature marginals, and let*

$$\mathcal{H} = \left\{\mathrm{sign}\left(w^\top x - \theta\right) \mid (w, \theta) \in \mathbb{R}^{n+1}\right\}$$

denote the class of halfspaces. There exists a degree $d = O(\varepsilon^{-4})$ such that running Algorithm 1 with ERM over parities of degree at most d produces a classifier $\overline{h}$ satisfying

$$l_{\mathcal{D}}\left(\overline{h}\right) \leq \min_{h \in \mathcal{H}} l_{\mathcal{D}}(h) + \varepsilon,$$

while using only $\mathcal{O}\left(\varepsilon^{-6}\, n^{40\,\varepsilon^{-4}}\right)$ labeled samples in $n^{\mathrm{poly}(1/\varepsilon)}$ time.

6.2. Boosting for reinforcement learning

The construction of near-optimal policy for reinforcement learning (RL) via boosting was first pursued in Brukhim et al. (2022). Ghai & Singh (2024) improve on these results. Modifying the RL setting to include the ability to sample trajectories *without observing reward*, we can apply our results to reduce the number of samples that require reward feedback. Such a feedback model could be useful where rollouts are cheap but reward feedback is not because it comes from human labeling or an expensive processes (Finn et al., 2016). In Appendix E, we provide a formal description of this modified setting.

Plugging Algorithm 1 into a modified meta-algorithm of Brukhim et al. (2022) in a manner that allows for trajectories without reward yields the following result for binary-action MDPs. Our result improves upon Ghai & Singh (2024) in that it requires fewer with-reward episodes. 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 E for a complete statement of results and details of the setting.

Theorem 6.2 (Informal; stated formally in Theorem E.1)**.** *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 $V^* - V^{\overline{\pi}} \leq (C_\infty\mathcal{E})/(1-\beta) + \varepsilon$, while sampling $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^4)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$ without reward feedback and $\mathcal{O}((\log|\mathcal{B}|)/\gamma^2\varepsilon^3)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$ with reward feedback. In the ν-*reset *access model, there is a setting of parameters such that Algorithm 4 when given access to $\mathcal{W}$ produces a policy $\overline{\pi}$ such that with probability $1 - \delta$, we have $V^* - V^{\overline{\pi}} \leq (D_\infty\mathcal{E}_\nu)/(1-\beta)^2 + \varepsilon$, while sampling $\mathcal{O}((\log|\mathcal{B}|)/\gamma^3\varepsilon^5)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$ without reward feedback and $\mathcal{O}((\log|\mathcal{B}|)/\gamma^2\varepsilon^4)$ episodes of length $\mathcal{O}((1-\beta)^{-1})$ with reward feedback.*

Dataset	No Added Noise		5% Noise		10% Noise		20% Noise	
	PAB	Ours	PAB	Ours	PAB	Ours	PAB	Ours
Ionosphere	0.87 ± 0.05	$\mathbf{0.91 \pm 0.04}$	0.88 ± 0.05	$\mathbf{0.90 \pm 0.04}$	0.84 ± 0.06	$\mathbf{0.90 \pm 0.04}$	0.81 ± 0.06	$\mathbf{0.83 \pm 0.06}$
Diabetes	0.84 ± 0.09	$\mathbf{0.89 \pm 0.07}$	$\mathbf{0.86 \pm 0.08}$	$\mathbf{0.86 \pm 0.09}$	0.79 ± 0.09	0.79 ± 0.10	0.76 ± 0.10	0.80 ± 0.10
Spambase	0.91 ± 0.02	$\mathbf{0.94 \pm 0.02}$	0.90 ± 0.03	$\mathbf{0.92 \pm 0.03}$	0.89 ± 0.03	$\mathbf{0.90 \pm 0.02}$	0.83 ± 0.04	$\mathbf{0.87 \pm 0.03}$
German	0.79 ± 0.07	$\mathbf{0.86 \pm 0.07}$	$\mathbf{0.84 \pm 0.08}$	$\mathbf{0.84 \pm 0.08}$	0.76 ± 0.08	$\mathbf{0.87 \pm 0.07}$	0.75 ± 0.08	$\mathbf{0.77 \pm 0.08}$
Sonar	0.78 ± 0.08	$\mathbf{0.92 \pm 0.05}$	0.68 ± 0.09	$\mathbf{0.89 \pm 0.06}$	0.84 ± 0.08	$\mathbf{0.87 \pm 0.07}$	0.69 ± 0.10	$\mathbf{0.77 \pm 0.08}$
Waveform	0.89 ± 0.02	$\mathbf{0.89 \pm 0.02}$	$\mathbf{0.88 \pm 0.03}$	0.87 ± 0.03	$\mathbf{0.86 \pm 0.03}$	0.86 ± 0.03	0.83 ± 0.03	$\mathbf{0.83 \pm 0.03}$
Average	0.84	**0.89**	0.84	**0.88**	0.81	**0.84**	0.78	**0.81**

Table 1. 50-fold cross-validated accuracies of the Potential based Agnostic Booster (PAB) (Kanade & Kalai, 2009) and our proposed boosting algorithm on six datasets with 0%, 5%, 10%, and 20% added label noise (during training). Sonar and Ionosphere have 50% of labels dropped while the remaining datasets have 90% of labels dropped. A final row is included for the average accuracy (evenly weighted) over all 6 datasets.

7. Experiments

In this section, we demonstrate the empirical viability of our approach. Table 1 showcases the results from our initial experiments comparing Algorithm 1 with the agnostic boosting method introduced by Kanade & Kalai (2009), herein referred to as the Potential-based Agnostic Booster (PAB). These evaluations were performed on various UCI classification datasets (Sigillito et al., 1989; Hopkins et al., 1999; Smith et al., 1988; Hofmann, 1994; Sejnowski & Gorman, 1988; Breiman & Stone, 1984), employing decision stumps (Pedregosa et al., 2011) as the weak learners. Notably, Algorithm 1 extends PAB to handle unlabeled data by incorporating our newly defined potential function, as defined in Equation (1).

To evaluate the robustness of all algorithms against label noise, we introduced noise levels of 5%, 10%, and 20% during the training phase. We randomly remove a certain percentage of labels from each dataset to create scenarios with both labeled and unlabeled instances. Specifically, we omitted 50% of labels for smaller datasets (Sonar and Ionosphere) and 90% for the other datasets. Our findings indicate that incorporating unlabeled examples leads to improved performance. This enhancement is likely attributed to the limitation of PAB in reusing samples, which consequently restricts the number of boosting iterations when the sample size is constrained. For a comprehensive overview of the experimental setup, please refer to Appendix C.

8. Conclusion

This paper aims to leverage unlabeled data to reduce the sample complexity of agnostic boosting. The theoretical improvements are stark. When given as much *unlabeled* data as the amount of *labeled* data required for existing approaches, the resultant sample complexity reduces to that of ERM, becoming essentially optimal. This is accomplished by a novel decomposable potential function, whose derivative naturally splits into two parts that are estimable independently by labeled and unlabeled data, respectively.

We end with a few concrete directions for future work. The possibility of achieving an optimal sample complexity for agnostic boosting in polynomial time without any concession remains open. In our view, an equally important and likely fruitful direction is to improve the oracle complexity of agnostic boosting where known results scale as $1/\varepsilon^2$, which is substantially worse than $\log 1/\varepsilon$ for the realizable case. Developing algorithms that adapt, on a per-round basis, to the weak learning edge γ_t is a key unresolved step to making agnostic boosting practical. Finally, it would be worthwhile to extend the sample complexity improvements in recent works to the *filtering* framework (cf. sub-sampling; see, for example, the discussion in Domingo (2000)), which essentially treats the weak learners as black-box learning algorithms and hence avoids appeals to uniform convergence arguments on the weak hypothesis class.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Alon, N., Gonen, A., Hazan, E., and Moran, S. Boosting simple learners. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 481–489, 2021.

Anthony, M. and Bartlett, P. L. *Neural network learning: Theoretical foundations*. cambridge university press, 2009.

Bartlett, P., Freund, Y., Lee, W. S., and Schapire, R. E. Boosting the margin: A new explanation for the effectiveness of voting methods. *The annals of statistics*, 26(5): 1651–1686, 1998.

Ben-David, S., Long, P. M., and Mansour, Y. 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*, pp. 507–516. Springer, 2001.

Beygelzimer, A., Kale, S., and Luo, H. Optimal and adaptive algorithms for online boosting. In *International Conference on Machine Learning*, pp. 2323–2331. PMLR, 2015.

Blanc, G., Hayderi, A., Koch, C., and Tan, L.-Y. The sample complexity of smooth boosting and the tightness of the hardcore theorem. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1431–1450. IEEE, 2024.

Breiman, L. and Stone, C. Waveform Database Generator (Version 1). UCI Machine Learning Repository, 1984. DOI: https://doi.org/10.24432/C5CS3C.

Brukhim, N., Chen, X., Hazan, E., and Moran, S. Online agnostic boosting via regret minimization. *Advances in Neural Information Processing Systems*, 33:644–654, 2020.

Brukhim, N., Hazan, E., and Singh, K. A boosting approach to reinforcement learning. *Advances in Neural Information Processing Systems*, 35:33806–33817, 2022.

Chapelle, O., Scholkopf, B., and Zien, A. Semi-supervised learning. 2006. *Cambridge, Massachusettes: The MIT Press View Article*, 2:1, 2006.

Chen, S.-T., Lin, H.-T., and Lu, C.-J. An online boosting algorithm with theoretical justifications. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1873–1880, 2012.

Chen, S.-T., Balcan, M.-F., and Chau, D. H. Communication efficient distributed agnostic boosting. In *Artificial Intelligence and Statistics*, pp. 1299–1307. PMLR, 2016.

da Cunha, A., Høgsgaard, M. M., Paudice, A., and Sun, Y. Revisiting agnostic boosting. *arXiv preprint arXiv:2503.09384*, 2025.

Diakonikolas, I., Impagliazzo, R., Kane, D. M., Lei, R., Sorrell, J., and Tzamos, C. Boosting in the presence of massart noise. In *Conference on Learning Theory*, pp. 1585–1644. PMLR, 2021.

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

Eisenstat, D. and Angluin, D. The vc dimension of k-fold union. *Information Processing Letters*, 101(5):181–184, 2007.

Feldman, V. Distribution-specific agnostic boosting. *Innovations in Theoretical Computer Science (ITCS)*, pp. 241–250, 2010.

Finn, C., Yu, T., Fu, J., Abbeel, P., and Levine, S. Generalizing skills with semi-supervised reinforcement learning. *arXiv preprint arXiv:1612.00429*, 2016.

Freund, Y. Boosting a weak learning algorithm by majority. *Information and computation*, 121(2):256–285, 1995.

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

Ghai, U. and Singh, K. Sample-efficient agnostic boosting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=ufKBRvYxtp.

Green Larsen, K. and Ritzert, M. Optimal weak to strong learning. *Advances in Neural Information Processing Systems*, 35:32830–32841, 2022.

Hazan, E. and Singh, K. Boosting for online convex optimization. In *International Conference on Machine Learning*, pp. 4140–4149. PMLR, 2021.

Hofmann, H. Statlog (German Credit Data). UCI Machine Learning Repository, 1994. DOI: https://doi.org/10.24432/C5NC77.

Hopkins, M., Reeber, E., Forman, G., and Suermondt, J. Spambase. UCI Machine Learning Repository, 1999. DOI: https://doi.org/10.24432/C53G6X.

Hopkins, M., Kane, D. M., Lovett, S., and Mahajan, G. Realizable learning is all you need. In *Conference on Learning Theory*, pp. 3015–3069. PMLR, 2022.

Huber, P. J. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pp. 492–518. Springer, 1992.

Jung, Y. H., Goetz, J., and Tewari, A. Online multiclass boosting. *Advances in neural information processing systems*, 30, 2017.

Kalai, A. and Servedio, R. A. Boosting in the presence of noise. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pp. 195–205, 2003.

Kalai, A. T., Mansour, Y., and Verbin, E. On agnostic boosting and parity learning. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 629–638, 2008.

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

Kanade, V. and Kalai, A. Potential-based agnostic boosting. *Advances in neural information processing systems*, 22, 2009.

Kearns, M. Learning boolean formulae or finite automata is as hard as factoring. *Technical Report TR-14-88 Harvard University Aikem Computation Laboratory*, 1988.

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

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

Kothari, P. K. and Livni, R. Improper learning by refuting. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Long, P. M. and Servedio, R. A. Random classification noise defeats all convex potential boosters. In *Proceedings of the 25th international conference on Machine learning*, pp. 608–615, 2008.

Lyu, X., Wu, H., and Yang, J. The cost of parallelizing boosting. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 3140–3155. SIAM, 2024.

Mansour, Y. and McAllester, D. Boosting using branching programs. *Journal of Computer and System Sciences*, 64 (1):103–112, 2002.

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

Raman, V. and Tewari, A. Online agnostic multiclass boosting. *Advances in Neural Information Processing Systems*, 35:25908–25920, 2022.

Schapire, R. E. The strength of weak learnability. *Machine learning*, 5:197–227, 1990.

Schapire, R. E. and Freund, Y. Boosting: Foundations and algorithms. *Kybernetes*, 42(1):164–166, 2013.

Schapire, R. E. and Singer, Y. Improved boosting algorithms using confidence-rated predictions. In *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 80–91, 1998.

Sejnowski, T. and Gorman, R. Connectionist Bench (Sonar, Mines vs. Rocks). UCI Machine Learning Repository, 1988. DOI: https://doi.org/10.24432/C5T01Q.

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

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

Smith, J. W., Everhart, J. E., Dickson, W., Knowler, W. C., and Johannes, R. S. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the annual symposium on computer application in medical care*, pp. 261. American Medical Informatics Association, 1988.

Talagrand, M. Sharper bounds for gaussian and empirical processes. *The Annals of Probability*, pp. 28–76, 1994.

Appendix

Map of the appendix. In Appendix A, we complete the proofs for results concerning improved unlabeled sample efficiency. Appendix B discusses the proofs for covariate shift. In Appendix C are provided experimental details not found in the main paper. Appendix D and Appendix E provide further details on applications of boosting to learning halfspace and reinforcement learning, respectively.

A. Improving unlabeled sample efficiency

A.1. Proofs for the result

Notice that in this section we use different choices of ϕ and ψ, those stated in Equation (4). However, Lemma 3.5 continues to hold. In fact, the latter only requires that $z\psi'(z) \geq 0$ for all z. To maintain the continuity of presentation, our organization and notation closely mirror Ghai & Singh (2024). Throughout this section, we will always set $\sigma = \eta/\gamma$.

Define $\Psi_{\mathcal{D}}(H) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\Psi(H(x))]$ and $\Psi'_{\mathcal{D}}(H,h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\Psi'(H(x))h(x)]$.

Proof of Theorem 4.1. This proof is almost identical to that of Theorem 4 in Ghai & Singh (2024). We reproduce it for completeness. In fact, the only change stems from the new upper bound on ε_{Gen} in Lemma A.1, which unlike the previous work makes a distinction between labeled and unlabeled samples.

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

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

But before that let us dispense with the random events at once. The success of Lemma A.1, the event that $\max_{t\in[T]} |\,\mathrm{corr}_{\widehat{D}_0}(\mathrm{sign}\, H_t) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}\, H_t)| \leq \varepsilon''/10$, and $\max_{t\in[T]} |\,\mathrm{corr}_{\widehat{D}'_t}(W_t) - \mathrm{corr}_{\mathcal{D}_t}(W_t)| \leq \varepsilon'/10$ can be ensured with probability $1-\delta$ by a simple application of Hoeffing's inequality and union bound given the setting of m and S_0 appropriately. We will soon set precise values of ε' and ε''.

Recall that $h_t = \eta(H_{t+1} - H_t)$. Equation (2) can be rearranged to get

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

where we use the fact that $\phi(0,y) - \min_z \phi(z,y) \leq 1$.

Case A: If $h_t = W_t/\gamma$, observe that $\mathrm{corr}_{\mathcal{D}_t}(W_t) \geq \mathrm{corr}_{D'_t}(W_t) - \varepsilon'/10 \geq \tau - \varepsilon'/10$. Now apply Lemma A.1 to get

$$-\Phi'_{\mathcal{D}}(H_t,h_t) \geq \frac{3}{\gamma}\mathrm{corr}_{\mathcal{D}_t}(W_t) - \frac{\varepsilon_{\text{Gen}}}{\gamma} \geq \frac{3}{\gamma}\left(\tau - \frac{\varepsilon'}{10}\right) - \frac{\varepsilon_{\text{Gen}}}{\gamma}$$

Case B: If $h_t = -\mathrm{sign}(H_t)$, then $\mathrm{corr}_{\mathcal{D}_t}(W_t) \leq \mathrm{corr}_{D'_t}(W_t) + \varepsilon'/10 \leq \tau + \varepsilon'/10$. Applying Lemma A.1, we get

$$3\left(\tau + \frac{\varepsilon'}{10}\right) \geq 3\mathrm{corr}_{\mathcal{D}_t}(W_t) \geq 3\gamma\mathrm{corr}_{\mathcal{D}_t}(h^*) - 3\varepsilon_0 \geq -\gamma\Phi'_{\mathcal{D}}(H_t,h^*) - 3\varepsilon_0 - \gamma\varepsilon_{\text{Gen}}$$

Using Lemma 3.5, this translates to

$$\begin{aligned}\Phi'_{\mathcal{D}}(H_t,-\mathrm{sign}(H_t)) &= -\Phi'_{\mathcal{D}}(H_t,\mathrm{sign}(H_t)) \leq -\Phi'_{\mathcal{D}}(H_t,h^*) - (\mathrm{corr}_{\mathcal{D}}(h^*) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}(H_t)))\\ &\leq -(\mathrm{corr}_{\mathcal{D}}(h^*) - \mathrm{corr}_{\mathcal{D}}(\mathrm{sign}(H_t))) + \frac{3}{\gamma}\left(\tau + \frac{\varepsilon'}{10} + \varepsilon_0\right) + \varepsilon_{\text{Gen}}.\end{aligned}$$

In either case, we have

$$-\Phi'_{\mathcal{D}}(H_t, h_t) \geq \min\left\{\frac{3}{\gamma}\left(\tau - \frac{\varepsilon'}{10}\right) - \frac{\varepsilon_{\text{Gen}}}{\gamma}, (\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\text{sign}(H_t))) - \frac{3}{\gamma}\left(\tau + \frac{\varepsilon'}{10} + \varepsilon_0\right) - \varepsilon_{\text{Gen}}\right\}.$$

Now, set

$$\tau = \frac{1}{3}\left(\frac{4}{\eta T} + \frac{\eta}{\gamma^2} + \frac{\varepsilon_{\text{Gen}}}{\gamma}\right)\gamma + \frac{\varepsilon'}{10}.$$

Hereafter let s be the time step satisfying

$$\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\text{sign}(H_s)) \leq \frac{3}{\gamma}(2\tau + \varepsilon_0) + \left(1 - \frac{1}{\gamma}\right)\varepsilon_{\text{Gen}} = \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \left(1 + \frac{1}{\gamma}\right)\varepsilon_{\text{Gen}} + 3\left(\frac{\varepsilon_0}{\gamma} + \frac{\varepsilon'}{5\gamma}\right).$$

Such a choice must exists, since otherwise we get for all t that

$$-\Phi'_{\mathcal{D}}(H_t, h_t) \geq \frac{3}{\gamma}\left(\tau - \frac{\varepsilon'}{10}\right) - \frac{\varepsilon_{\text{Gen}}}{\gamma} = \frac{4}{(\eta T)} + \frac{\eta}{\gamma^2},$$

which contradicts Equation (2). Combining this with the observation that $\overline{h}$ minimizes the correlation on $\widehat{D}_0$, we get

$$\text{corr}_{\mathcal{D}}(h^*) - \text{corr}_{\mathcal{D}}(\overline{h}) \leq \frac{8}{\eta T} + \frac{2\eta}{\gamma^2} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} + \frac{3}{\gamma}\left(\varepsilon_0 + \frac{\varepsilon'}{5}\right) + \frac{\varepsilon''}{5}.$$

Setting $\varepsilon' = \gamma\varepsilon/100$, $\varepsilon'' = \gamma\varepsilon/100$ and plugging in the proposed hyper-parameters with appropriate constants yields the claimed result. □

Proof of Lemma A.1. By the definition of $\mathcal{D}_t$, we have that

$$\text{corr}_{\mathcal{D}_t}(h) = \frac{1}{3}\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] + \frac{2}{3}\mathbb{E}_{\mathcal{D}'_t}[yh(x)] = \frac{1}{3}\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] + \frac{2}{3}\,\text{corr}_{\mathcal{D}'_t}(h),$$

Since $\widehat{D}$ is composed of IID draws from $\mathcal{D}$, the standard uniform convergence result via union bound gets that for some constant $C \geq 0$ we have with probability $1 - \delta$ that $|\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \text{corr}_{\mathcal{D}}(h)|$ is at most $\sqrt{(\log|\mathcal{B}| + \log\delta^{-1})/S}$.

By inspecting Line 6 in Algorithm 2 here and Line 5 in Algorithm 1 in Ghai & Singh (2024) with the substitution that $y = 1$, we note that the two are identical. Therefore, we can apply the following result from Ghai & Singh (2024).

Lemma A.2 (Lemma 6 in Ghai & Singh (2024))**.** *Setting $\sigma = \eta/\gamma$. There exists a universal constant $C > 0$ such that with probability $1 - \delta$, for all $t \in [T], h \in \mathcal{B} \cup \{h^*\}$:*

$$\left|\Psi'_{\mathcal{D}}(H_t, h) + 2\,\text{corr}_{\mathcal{D}'_t}(h)\right| \leq \frac{C\eta}{\gamma}\left(\sqrt{\frac{\log|\mathcal{B}|T/\delta}{\sigma U}} + \log|\mathcal{B}|T/\delta\right).$$

Since $\Phi_{\mathcal{D}}(H) = \Psi_{\mathcal{D}}(H) - \text{corr}_{\mathcal{D}}(H)$, we get

$$\left|\frac{1}{3}\Phi'_{\mathcal{D}}(H_t, h) + \text{corr}_{\mathcal{D}_t}(h)\right| \leq \frac{1}{3}|\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \text{corr}_{\mathcal{D}}(h)| + \frac{1}{3}\left|\Psi'_{\mathcal{D}}(H_t, h) + 2\,\text{corr}_{\mathcal{D}'_t}(h)\right|$$

$$\Phi'_{\mathcal{D}}(H_t, h) = \mathop{\mathbb{E}}_{x\sim\mathcal{D}_{\mathcal{X}}}[\psi'(H_t(x))h(x)] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yh(x)],$$

completing the proof. □

A.2. Trading off oracle Complexity and unlabeled sample complexity

Theorem A.3 (Main theorem with unlabeled data reuse)**.** *For any $\varepsilon, \delta > 0$, there is an instantiation of parameters for which Algorithm 2 guarantees with probability $1 - \delta - T\delta_0$ that*

$$\operatorname{corr}_{\mathcal{D}}(\overline{h}) \geq \max_{h \in \mathcal{H}} \operatorname{corr}_{\mathcal{D}}(h) - \frac{3\varepsilon_0}{\gamma} - \varepsilon.$$

During its execution, Algorithm 2 makes $\mathcal{O}(1/\gamma^2\varepsilon^2)$ calls to the weak learner, and samples $S + S_0 = \mathcal{O}(\log|\mathcal{B}|/\gamma^2\varepsilon^2)$ labeled samples and $TU = \mathcal{O}(\log|\mathcal{B}|/\gamma^3\varepsilon^3 + (\log|\mathcal{B}|)^3/\gamma^3\varepsilon^2)$ unlabeled samples.

Proof. The proof is identical to that of Theorem 4.1, except crucially to the substitution of the following bound on ε_{Gen}.

There exists a universal constant $C > 0$ such that with probability $1 - \delta$, for any $t \in [T], h \in \mathcal{B} \cup \{h^*\}$:

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

To prove this claim itself, we follow the same recipe as in the proof of Lemma A.1. Once again we observe that By Line 6 in Algorithm 2 here is identical to Line 5 in Algorithm 1 in Ghai & Singh (2024) with the substitution that $y = 1$. Therefore, we can apply the following result from Ghai & Singh (2024).

Lemma A.4 (Lemma 15 in Ghai & Singh (2024))**.** *Setting $\sigma = \eta/\gamma$. There exists a universal constant $C > 0$ such that with probability $1 - \delta$, for all $t \in [T], h \in \mathcal{B} \cup \{h^*\}$:*

$$\left|\Psi'_{\mathcal{D}}(H_t, h) + 2\operatorname{corr}_{\mathcal{D}'_t}(h)\right| \leq \frac{C\eta}{\gamma}\left(\sqrt{\frac{\log|\mathcal{B}|T/\delta}{\sigma U}} + \log|\mathcal{B}|T/\delta\right).$$

□

B. Resiliency against covariate shift

Let $\psi(z)$ be the Huber loss. Instead of definite a scalar potential, this time we will directly define the population potential measure that involves both $\mathcal{D}$ and $\mathcal{Q}$. Recall that $\mathcal{C}_{\mathcal{X}} \geq \|d\mathcal{D}_{\mathcal{X}}/d\mathcal{Q}\|$ has to be at least one.

$$\Phi(H) = \mathcal{C}_{\mathcal{X}} \mathop{\mathbb{E}}_{x\sim\mathcal{Q}}[\psi(H(x))] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yH(x)]$$
$$\Phi'(H, h) = \mathcal{C}_{\mathcal{X}} \mathop{\mathbb{E}}_{x\sim\mathcal{Q}}[\psi'(H(x))h(x)] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yh(x)]$$

We describe a variant of Algorithm 1 that can tolerate a mismatch between $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{Q}$. The key modification happens in Line 6 of Algorithm 3, where pseudo-labeled samples created from the unlabeled distribution $\mathcal{Q}$ are sampled a higher rate than labeled samples.

The uniform convergence still applies after some minor adjustment.

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

$$|\Phi'(H_t, h) + (1 + \mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\mathcal{D}_t}(h)| \leq \varepsilon_{Gen} \coloneqq C\left(\sqrt{\frac{\mathrm{VC}(\mathcal{B}) + \log\frac{1}{\delta}}{S}} + \mathcal{C}_{\mathcal{X}}\sqrt{\frac{\mathrm{VC}(\mathcal{B}) + \log\frac{1}{\delta}}{U}}\right).$$

Proof of Lemma B.1. By the definition of $\mathcal{D}_t$, we have that

$$\operatorname{corr}_{\mathcal{D}_t}(h) = \frac{1}{1 + \mathcal{C}_{\mathcal{X}}}\left(\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \mathcal{C}_{\mathcal{X}}\widehat{\mathbb{E}}_{\widehat{D}_t}[\psi'(H_t(x))h(x)]\right),$$

Algorithm 3 Covariate-shift Resistant Agnostic Boosting with Unlabeled Data

Inputs: Samplers for labeled data from $\mathcal{D}$ and unlabeled data from $\mathcal{Q}$, γ-agnostic weak learning oracle $\mathcal{W}$, parameters $\eta, T, \tau, S, U, S_0, m$.
Initialize a zero hypothesis $H_1 = \mathbf{0}$.
Sample S *labeled* examples to create dataset $\widehat{D}$.
for $t = 1$ to T **do**
  Sample U *unlabeled* examples to create dataset $\widehat{D}_t$.
  Construct a resampling distribution $\mathcal{D}_t$ that:
    A. With probability $\frac{1}{\mathcal{C}_\mathcal{X}}$, returns (x, y) picked uniformly from $\widehat{D}$.
    B. With remaining probability, picks x uniformly from $\widehat{D}_t$, and returns $(x, \widehat{y})$, where $\widehat{y}$ is chosen as

$$\widehat{y} = \begin{cases} +1 & \text{with probability } p_t(x) = \frac{1-\psi'(H_t(x))}{2}, \\ -1 & \text{with remaining probability.} \end{cases}$$

  Sample m times from $\mathcal{D}_t$ to create dataset $\widehat{D}'_t$.
  Call the weak learner on $\widehat{D}'_t$ to get $W_t = \mathcal{W}(\widehat{D}'_t)$.
  if $\operatorname{corr}_{\widehat{D}'_t}(W_t) = \sum_{(x,\widehat{y})\in\widehat{D}'_t} \widehat{y} W_t(x) > \tau$ **then**
    Update $H_{t+1} = H_t + \eta W_t/\gamma$.
  else
    Update $H_{t+1} = H_t - \eta \operatorname{sign}(H_t)_t$.
  end if
end for
Sample S_0 *labeled* examples to create dataset $\widehat{D}_0$.
Output $\overline{h} = \underset{h\in\{\operatorname{sign}(H_t):t\in[T]\}}{\arg\max} \sum_{(x,y)\in\widehat{D}_0} yh(x)$.

where we use the fact that Line 6.B in Algorithm 3 ensures $\mathbb{E}[\widehat{y}|x] = -\psi'(H_t(x))$. Since $\widehat{D}$ and $\widehat{D}_t$ are composed of IID draws from $\mathcal{D}$ and $\mathcal{Q}$ respectively, we have that, for some constant $C \geq 0$ we have with probability $1-\delta$ that $|\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \operatorname{corr}_\mathcal{D}(h)| \leq \sqrt{(\mathrm{VC}(\mathcal{B}) + \log\delta^{-1})/S}$ and $|\widehat{\mathbb{E}}_{\widehat{D}_t}[\psi'(H_t(x))h(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\psi'(H_t(x))h(x)]| \leq \sqrt{(\mathrm{VC}(\mathcal{B}) + \log\delta^{-1})/U}$.

$$|(1+\mathcal{C}_\mathcal{X})\operatorname{corr}_{\mathcal{D}_t}(h) + \Phi'(H_t, h)| \leq \mathcal{C}_\mathcal{X}|\widehat{\mathbb{E}}_{\widehat{D}_t}[\psi'(H_t(x))h(x)] - \mathbb{E}_{x\sim\mathcal{Q}}[\psi'(H_t(x))h(x)]| + |\widehat{\mathbb{E}}_{\widehat{D}}[yh(x)] - \operatorname{corr}_\mathcal{D}(h)|$$

The decomposition above completes the proof. □

We will need the following well-known property of Random-Nikodym derivatives.

Lemma B.2. *For any non-negative function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$, it is true that $C_\mathcal{X}\,\mathbb{E}_{x\sim Q}[f(x)] \geq \mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}}[f(x)]$.*

Proof. Since $\mathcal{C}_\mathcal{X} \geq \|d\mathcal{D}_\mathcal{X}/d\mathcal{Q}\|$, we have $C_\mathcal{X}\,\mathbb{E}_{x\sim Q}[f(x)] \geq \mathbb{E}_{x\sim Q}\left[f(x)\frac{d\mathcal{D}_\mathcal{X}}{d\mathcal{Q}}(x)\right] = \mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}}[f(x)]$. □

The key idea in the analysis occurs in the following lemma. Essentially, it says oversampling the ψ part from unlabeled data does not hurt the correlation gap.

Lemma B.3. *For any real-valued classifier $H : \mathcal{X} \to \mathbb{R}$, we have*

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

Proof. Using Lemma B.2 below, we arrive at

$$\begin{aligned}
\Phi'(H, \operatorname{sign}(H)) - \Phi'(H, h^*) &= \mathcal{C}_\mathcal{X} \mathop{\mathbb{E}}_{x\sim\mathcal{Q}}[\psi'(H(x))(\operatorname{sign}(H(x)) - h^*(x))] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[y(\operatorname{sign}(H(x)) - h^*(x))] \\
&\geq \mathop{\mathbb{E}}_{x\sim\mathcal{D}_\mathcal{X}}[\psi'(H(x))(\operatorname{sign}(H(x)) - h^*(x))] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[y(\operatorname{sign}(H(x)) - h^*(x))] \\
&= \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[(\psi'(H(x)) - y)(\operatorname{sign}(H(x)) - h^*(x))]\,.
\end{aligned}$$

Recall that z and $\psi'(z)$ always have the same sign, and hence so do $\psi'(z)$ and $\operatorname{sign}(z)$. This ensures non-negativity as $\psi'(H(x))(\operatorname{sign}(H(x)) - h^*(x)) = |\psi'(H(x))| - \psi'(H(x))h^*(x)$, since $h^*(x)$ is restricted to $\{\pm 1\}$.

From here onward, our original proof strategy work. In particular since $y^2 = 1$, we get

$$\Phi'(H, \operatorname{sign}(H)) - \Phi'(H, h^*) \geq \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}} \left[(1 - y\psi'(H(x)))y(\operatorname{sign}(H(x)) - h^*(x))\right].$$

As before, consider any (x, y) such that $yH(x) > 0$: Here $y(h^*(x) - \operatorname{sign}(H(x))) < 0$. Furthermore, since y and $H(x)$ have the same sign, so do y and $\psi'(H(x))$, and hence $(1 - y\psi'(H(x))) \leq 1$. Similarly, whenever $yH(x) < 0$: Then $y(h^*(x) - \operatorname{sign}(H(x))) > 0$, and y and $\psi'(H(x))$ have opposite signs that imply $(1 - y\psi'(H(x))) \geq 1$.

Now the claim follows as

$$\begin{aligned}
&\Phi'_{\mathcal{D}}(H, \operatorname{sign}(H)) - \Phi'_{\mathcal{D}}(H, h^*)) \\
&= \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{\mathcal{X}}} \left[\mathbb{1}_{yH(x)\geq 0} \underbrace{(1 - y\psi'(H(x)))}_{\leq 1} \underbrace{y(h^*(x) - \operatorname{sign}(H(x)))}_{\leq 0} + \mathbb{1}_{yH(x)<0} \underbrace{(1 - y\psi'(H(x)))}_{\geq 1} \underbrace{y(h^*(x) - \operatorname{sign}(H(x)))}_{\geq 0} \right] \\
&\geq \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}} [y(h^*(x) - \operatorname{sign} H(x))] \\
&= \operatorname{corr}_{\mathcal{D}}(h^*) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign}(H)).
\end{aligned}$$

□

We are finally ready to prove the main result.

Proof of Theorem 3.1. Let us dispense with the random events at once. The success of Lemma B.1, the event that $\max_{t\in[T]} |\operatorname{corr}_{\widehat{D}_0}(\operatorname{sign} H_t) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign} H_t)| \leq \varepsilon/10$, and $\max_{t\in[T]} |\operatorname{corr}_{\widehat{D}'_t}(W_t) - \operatorname{corr}_{\mathcal{D}_t}(W_t)| \leq \gamma\varepsilon/20\mathcal{C}_{\mathcal{X}}$ can be ensured with probability $1 - \delta$ by a simple application of Hoeffing's inequality and union bound given the setting of $m = m(\varepsilon_0, \delta_0) + \mathcal{O}((\mathcal{C}_{\mathcal{X}})^2/\varepsilon^2\gamma^2)$ and $S_0 = \mathcal{O}(1/\gamma^2\varepsilon^2)$. Similarly, $\varepsilon_{\text{Gen}} \leq \gamma\varepsilon/10$ holds in Lemma 3.3 for $S = \Omega((\mathrm{VC}(\mathcal{B}) + \log\delta^{-1})/\gamma^2\varepsilon^2)$ and $U = \Omega((\mathcal{C}_{\mathcal{X}})^2(\mathrm{VC}(\mathcal{B}) + \log\delta^{-1})/\gamma^2\varepsilon^2)$.

Let $h_t = (H_{t+1} - H_t)/\eta$. Since ψ is 1-smooth, we have

$$\Phi_{\mathcal{D}}(H_{t+1}) - \Phi_{\mathcal{D}}(H_t) \leq \eta\Phi'_{\mathcal{D}}(H_t, h_t) + \frac{\eta^2\mathcal{C}_{\mathcal{X}}}{2\gamma^2}. \tag{5}$$

Case A: Consider any step t where $\operatorname{corr}_{\widehat{D}'_t}(W_t) > \tau$. Here $h_t = W_t/\gamma$. It follows from Lemma B.1 that

$$\begin{aligned}
\Phi'(H_t, h_t) &\leq -\frac{(1 + \mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\mathcal{D}_t}(h_t)}{\gamma} + \frac{\varepsilon_{\text{Gen}}}{\gamma} \\
&\leq -\frac{(1 + \mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\widehat{D}'_t}(h_t)}{\gamma} + \frac{2\varepsilon_{\text{Gen}}}{\gamma} + \frac{\varepsilon}{5} \\
&\leq -\varepsilon
\end{aligned}$$

where $\tau = 2\gamma\varepsilon/(1 + \mathcal{C}_{\mathcal{X}})$ and $\eta = \gamma^2\varepsilon/\mathcal{C}_{\mathcal{X}}$. By Equation (5), the potential drops as $\Phi(H_{t+1}) - \Phi(H_t) \leq -\gamma^2\varepsilon^2/2\mathcal{C}_{\mathcal{X}}$.

Case B: Consider any step t where $\operatorname{corr}_{\widehat{D}'_t}(W_t) \leq \tau$ and *crucially* $\Phi(H_t, \operatorname{sign} H_t) \geq \varepsilon$. Here $h_t = -\operatorname{sign} H_t$. Since $\Phi'(H_t, h_t) = -\Phi'(H_t, \operatorname{sign} H_t) \leq -\varepsilon$, by Equation (5), we have $\Phi(H_{t+1}) - \Phi(H_t) \leq -\gamma^2\varepsilon^2/2\mathcal{C}_{\mathcal{X}}$.

At initialization, $\Phi(\mathbf{0}) = 0$. Further for any $H : \mathcal{X} \to \mathbb{R}$, using non-negativity of ψ, we have

$$\Phi(H) = \mathcal{C}_{\mathcal{X}} \mathop{\mathbb{E}}_{x\sim\mathcal{Q}}[\psi(H(x))] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yH(x)] \geq \mathop{\mathbb{E}}_{x\sim\mathcal{D}_{\mathcal{X}}}[\psi(H(x))] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[yH(x)] \geq \frac{1}{2}.$$

Thus, at initialization Φ is at most half away from its minimum. Thus, setting $T = 2\mathcal{C}_{\mathcal{X}}/\gamma^2\varepsilon^2$, there must arise an iterate such that neither Case A nor Case B hold. That is, there is some $s \in [T]$ such that $\operatorname{corr}_{\widehat{D}'_s}(W_s) \leq \tau$ and $\Phi_{\mathcal{D}}(H_s, \operatorname{sign} H_s) \leq \varepsilon$.

Now using Lemma B.1 and that the weak learner γ-approximately maximizes correlation (Definition 2.1), we have

$$\begin{aligned}
-\Phi'(H_s, h^*) &\le (1+\mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\mathcal{D}_s}(h^*) + 2\varepsilon_{\text{Gen}} \\
&\le \frac{(1+\mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\mathcal{D}_s}(W_s)}{\gamma} + \frac{(1+\mathcal{C}_{\mathcal{X}})\varepsilon_0}{\gamma} + 2\varepsilon_{\text{Gen}} \\
&\le \frac{(1+\mathcal{C}_{\mathcal{X}})\operatorname{corr}_{\widehat{D}'_s}(W_s)}{\gamma} + \frac{(1+\mathcal{C}_{\mathcal{X}})\varepsilon_0}{\gamma} + \frac{\varepsilon}{5} \\
&\le \frac{(1+\mathcal{C}_{\mathcal{X}})\varepsilon_0}{\gamma} + \frac{12\varepsilon}{5}
\end{aligned}$$

where in the last line we recall $\tau = 2\varepsilon\gamma/(1+\mathcal{C}_{\mathcal{X}})$ and $\varepsilon_{\text{Gen}} = \varepsilon\gamma/10$.

By Lemma 3.5, we have

$$\begin{aligned}
\operatorname{corr}_{\mathcal{D}}(h^*) - \operatorname{corr}_{\mathcal{D}}(\operatorname{sign}(H_s)) &\le \Phi'_{\mathcal{D}}(H_s, \operatorname{sign}(H_s)) - \Phi'_{\mathcal{D}}(H_s, h^*) \\
&\le \frac{(1+\mathcal{C}_{\mathcal{X}})\varepsilon_0}{\gamma} + \frac{17\varepsilon}{5}.
\end{aligned}$$

To complete the proof, we observe that

$$\begin{aligned}
\operatorname{corr}_{\mathcal{D}}(\overline{h}) &\le \operatorname{corr}_{\widehat{D}_0}(\overline{h}) + \varepsilon/10 \\
&\le \operatorname{corr}_{\widehat{D}_0}(\operatorname{sign} H_s) + \varepsilon/10 \\
&\le \operatorname{corr}_{\mathcal{D}}(\operatorname{sign} H_s) + \varepsilon/5 \\
&\le (1+\mathcal{C}_{\mathcal{X}})\varepsilon_0/\gamma + 18\varepsilon/5,
\end{aligned}$$

where we use the fact that $\overline{h}$ maximizes the empirical correlation on the dataset $\widehat{D}_0$. □

C. Additional experimental details

For PAB, the number of samples that can be fed to a week learner in a round scales inversely with the number of boosting rounds, as the algorithm requires fresh samples each round.As such, we perform a grid search on the number of boosting rounds with $T \in \{25, 50, 100\}$, while we just use 100 for our implementation of Algorithm 1. In both algorithms we search over the parameter m, the number of samples we feed to the weak learner each round with a grid of $\{5, 20, 50, 100\}$, though if such a setting is invalid for PAB, we continue until all samples are used.

Our experiments were performed using the fractional relabeling scheme stated in (Kanade & Kalai, 2009), 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 equal to their sampling probabilities. Experiments are all run on an M1 Macbook Pro and complete within an hour. Multiclass datasets are converted to binary.

D. Proof of Theorem 6.1

Proof. 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 (Klivans et al., 2004). As such, an n-dimensional halfspace can be approximated with uniform weighting on the hypercube to ε^2 ℓ_2-error using degree-limited $\mathcal{B}_{n,d} = \{\pm\chi_S : |S| \le d\}$ 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 D.1 (Lemma 5 in (Kalai et al., 2008))**.** *Let $\mathcal{D}$ be any data distribution over $\{\pm 1\}^n \times \{-1, 1\}$ with marginal distribution Unif$(\{\pm 1\}^n)$ on the features. For any fixed ε and $d = 20\varepsilon^{-4}$, there exists some $h \in \mathcal{B}_{n,d}$ such*

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

The result follows directly from the preceding lemma, which provides a weak learner for the task, and Theorem 3.1. We note that $|\mathcal{B}_{n,d}| < n^d$ and $\gamma = n^{-d}$, so

$$\frac{\log|\mathcal{B}_{n,d}|}{\gamma^2\varepsilon^2} \le \frac{dn^{2d}\log(n)}{\varepsilon^2}.$$

The unlabeled samples used in Algorithm 1 can be produced by sampling from the hypercube adding to the $n^{\mathrm{poly}(1/\varepsilon)}$ runtime, but not the sample complexity. □

E. Boosting for reinforcement learning

In this section, we consider boosting in the reinforcement learning setting. We wish to separately consider the number of reward-annotated episodes against the number of reward-free episodes needed to learn a near-optimal policy.

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 (which is sometimes available), $P : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ 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. Let $Q^\pi(s, a)$ and $V^\pi(s)$ be the state-action and state value functions. Let $V^\pi_\mu = \mathbb{E}_{s\sim\mu}[V(s)]$ be the expected total reward when starting from the start state distribution μ, and we will say $V^\pi_{\mu_0} = V^\pi$. 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.

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. In this setting, we do not always assume the reward is revealed. We consider a model where we can rollout a policy and observe rewards or alternatively can just observe the state trajectories.

Since we have binary actions, our weak learners are policies, which we denote π instead of h. This notion is equivalent to that used by Brukhim et al. (2022) and Ghai & Singh (2024), because for binary actions, a random policy induces an accuracy of half regardless of the distribution over features and labels.

Say $\pi^* \in \arg\max_\pi V^\pi$ 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 E.1. *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 4 when given access to $\mathcal{W}$ produces a policy $\overline{\pi}$ such that with probability $1 - \delta$, we have*

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

while sampling

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

episodes of length $\mathcal{O}((1 - \beta)^{-1})$ without reward feedback (via Algorithm 6) and

$$\mathcal{O}\left(\frac{C_\infty^4 \log|\mathcal{B}|}{(1 - \beta)^7 \gamma^2 \varepsilon^3}\right)$$

episodes of length $\mathcal{O}((1 - \beta)^{-1})$ with reward feedback (via Algorithm 5).

Algorithm 4 RL Boosting adapted from (Brukhim et al., 2022)

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 2 to get π'_t, using

- Algorithm 5 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 as the labeled samples.
- Algorithm 6 to produce a distribution over states by executing the current policy π_{t-1} starting from the initial state distribution μ as the unlabeled samples.

  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 $t' = \arg\max_t \widehat{V^{\pi_t}}$.

Algorithm 5 Trajectory Sampler adapted from (Brukhim et al., 2022)

Sample state $s_0 \sim \mu$ and action $a' \sim \text{Unif}(\mathcal{A})$.
Sample $s \sim \mu^\pi$ as follows: at every step h, with probability β, execute π; else, accept s_h.
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.
Define the vector $\widehat{Q}$, such that for all $a \in A$, $\widehat{Q}(a) = 2R(s_h, a') \cdot \mathbb{I}_{a=a'}$.
With probability $C\widehat{Q}(a')$, set $y = a'$ else set $y \in \mathcal{A} - \{a'\}$, where $C = (1 - \beta)/2$.
return $(s_h, \widehat{Q}, y)$.

*In the ν-*reset *access model, there is a setting of parameters such that Algorithm 4 when given access to $\mathcal{W}$ produces a policy $\overline{\pi}$ such that with probability $1 - \delta$, we have*

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

while sampling

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

episodes of length $\mathcal{O}((1-\beta)^{-1})$ without reward feedback (via Algorithm 6) and

$$\mathcal{O}\left(\frac{D_\infty^4 \log|\mathcal{B}|}{(1-\beta)^{12}\gamma^2\varepsilon^4}\right)$$

episodes of length $\mathcal{O}((1-\beta)^{-1})$ with reward feedback (via Algorithm 5).

Proof. The proof follows by applying the result in Theorem 4.1 within the proof of Theorem 22 from Ghai & Singh (2024).

For the episodic model, applying the second part of Theorem 9 in (Brukhim et al., 2022), while noting the smoothness of V^π, and combining the result with Lemma 18 and Lemma 11 in (Brukhim et al., 2022), we have with probability $1 - T\delta$

Following the logic of Theorem 22 from Ghai & Singh (2024), we need to ensure is that output of Algorithm 2 as instantiated in Algorithm 4 every round has an excess correlation gap over the best policy Π no more that $(1-\beta)^2\varepsilon/C_\infty$, which Algorithm 2 assures us can be accomplished with $\mathcal{O}\left(\frac{C_\infty^3 \log|\mathcal{B}|}{(1-\beta)^6\gamma^3\varepsilon^3}\right)$ unlabeled samples and $\mathcal{O}\left(\frac{C_\infty^2 \log|\mathcal{B}|}{(1-\beta)^4\gamma^2\varepsilon^2}\right)$ labeled samples. The total number of samples is $T = \mathcal{O}\left(\frac{C_\infty^2}{(1-\beta)^3\varepsilon}\right)$ times greater.

Similarly, for the ν-reset model, we need to ensure is that output of Algorithm 2 as instantiated in Algorithm 4 every round has an excess correlation gap over the best policy Π no more that $(1-\beta)^3\varepsilon/D_\infty$, which Algorithm 2 assures us can be

Algorithm 6 Reward-free Trajectory Sampler

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 $1-\beta$.
4: **return** s_h.

accomplished with $\mathcal{O}\left(\frac{D_\infty^3 \log|\mathcal{B}|}{(1-\beta)^9\gamma^3\varepsilon^3}\right)$ unlabeled samples and $\mathcal{O}\left(\frac{D_\infty^2 \log|\mathcal{B}|}{(1-\beta)^6\gamma^2\varepsilon^2}\right)$ labeled samples. The total number of samples is $T = \mathcal{O}\left(\frac{D_\infty^2}{(1-\beta)^6\varepsilon^2}\right)$ times greater. □