Apple Tasting: Combinatorial Dimensions and Minimax Rates

Vinod Raman * VKRAMAN@UMICH.EDU
University of Michigan

Unique Subedi * SUBEDI@UMICH.EDU
University of Michigan

Ananth Raman ANANTHSRAMAN2007@GMAIL.COM
Bridgewater, New Jersey

Ambuj Tewari TEWARIA@UMICH.EDU
University of Michigan

Editors: Shipra Agrawal and Aaron Roth

Abstract

In online binary classification under *apple tasting* feedback, the learner only observes the true label if it predicts "1". First studied by Helmbold et al. (2000a), we revisit this classical partial-feedback setting and study online learnability from a combinatorial perspective. We show that the Littlestone dimension continues to provide a tight quantitative characterization of apple tasting in the agnostic setting, closing an open question posed by Helmbold et al. (2000a). In addition, we give a new combinatorial parameter, called the Effective width, that tightly quantifies the minimax expected number of mistakes in the realizable setting. As a corollary, we use the Effective width to establish a *trichotomy* of the minimax expected number of mistakes in the realizable setting. In particular, we show that in the realizable setting, the expected number of mistakes of any learner, under apple tasting feedback, can only be either $\Theta(1)$, $\Theta(\sqrt{T})$, or $\Theta(T)$. This is in contrast to the full-information realizable setting where only $\Theta(1)$ and $\Theta(T)$ are possible.

Keywords: Online Learning, Partial Feedback, Classification

1. Introduction

In the standard online binary classification setting, a learner plays a repeated game against an adversary. In each round, the adversary picks a labeled example $(x, y) \in \mathcal{X} \times \{0, 1\}$ and reveals the unlabeled example x to the learner. The learner observes x and then makes a prediction $\hat{y} \in \{0, 1\}$. Finally, the adversary reveals the true label y and the learner suffers the loss $\mathbb{1}\{\hat{y} \neq y\}$ (Littlestone, 1987). In many situations, receiving feedback after every prediction may not be realistic. For example, in spam filtering, emails that are classified as spam are often not verified by the user. Accordingly, the learner only receives feedback when an email is classified as "not spam." In recidivism prediction, a person whose is predicted to re-commit a crime may not be released. Accordingly, we will not know whether this person would have re-committed a crime had they been released. Situations like these are known formally as "apple tasting" (Helmbold et al., 2000a). In the generic model, a learner observes a sequence of apples, some of which may be rotten. For each apple, the learner must decide whether to discard or taste the apple. The learner suffers a loss if they discard a good apple or if they taste a rotten apple. Crucially, when the learner discards an apple, they do not receive any feedback on whether the apple was rotten or not.

* Equal contribution

Binary online classification under apple tasting feedback was first studied by Helmbold et al. (2000a) in the realizable setting. Here, they give a simple and generic conversion of a deterministic online learner in the full-information setting into a randomized online learner in the apple tasting setting. In particular, they show that if M_+ and M_- are upper bounds on the number of false positive and false negative mistakes of the deterministic online learner respectively, then the expected number of mistakes made by their conversion, under apple tasting feedback, is at most $M_+ + 2\sqrt{TM_-}$. Along with these upper bounds, they provide lower bounds on the expected number of mistakes for randomized apple tasting learners in terms of the number of false positive and false negative mistakes made by any deterministic online learner in the full-information setting. That is, if there exists $M_+, M_- \in \mathbb{N}$ such that every deterministic online learner in the full-information setting makes either at least M_+ false positive mistakes *or* M_- false negative mistakes, then every randomized online learner makes at least $\frac{1}{2}\min\left\{\frac{1}{2}\sqrt{TM_-}, M_+\right\}$ expected number of mistakes under apple tasting feedback. Finally, as an open question, they ask whether their results can be extended to the harder agnostic setting where the true labels can be noisy.

While Helmbold et al. (2000a) establish bounds on the minimax expected number of mistakes in the realizable setting, their bounds are in terms of the existence of an algorithm with certain properties. This is in contrast to much of online learning theory, where minimax regret is often quantified in terms of combinatorial dimensions that capture the complexity of the hypothesis class (Littlestone, 1987; Ben-David et al., 2009; Daniely et al., 2011; Rakhlin et al., 2015). Accordingly, we revisit apple tasting and study online learnability from a combinatorial perspective. In particular, we are interested in identifying combinatorial dimensions that tightly quantify the minimax regret for apple tasting in both the realizable and agnostic settings. To that end, our main contributions are:

(1) We close the open question posed by Helmbold et al. (2000a) by showing that the minimax expected regret in the agnostic setting, under apple tasting feedback, is at most $3\sqrt{\mathrm{L}(\mathcal{H})T\log(T)}$ and at least $\sqrt{\frac{\mathrm{L}(\mathcal{H})T}{8}}$, where $\mathrm{L}(\mathcal{H})$ is the Littlestone dimension of $\mathcal{H}$.

(2) On the other hand, we show that the Littlestone dimension alone does not give a tight quantitative characterization in the realizable setting. Instead, we show that the minimax expected number of mistakes in the realizable setting, under apple tasting feedback is

$$\Theta\Big(\max\Big\{\sqrt{(\mathrm{W}(\mathcal{H})-1)T}, 1\Big\}\Big),$$

where $\mathrm{W}(\mathcal{H})$ is the *Effective width* of $\mathcal{H}$, a new combinatorial parameter that accounts for the asymmetric nature of the feedback.

(3) Using the bound above, we establish the following trichotomy on the minimax rates in the realizable setting: (i) $\Theta(1)$ when $\mathrm{W}(\mathcal{H}) = 1$, (ii) $\Theta(\sqrt{T})$ when $1 < \mathrm{W}(\mathcal{H}) < \infty$, and (iii) $\Theta(T)$ when $\mathrm{W}(\mathcal{H}) = \infty$.

To prove (1), we extend the EXP3.G algorithm from Alon et al. (2015) to binary prediction with expert advice. Then, we use the standard technique from Ben-David et al. (2009) to construct an agnostic learner using a realizable, mistake-bound learner in the full-information setting. To prove the upper bound in (2), we define a new combinatorial parameter, called the Effective width, and use it to construct a deterministic online learner in the realizable, *full-information* feedback setting

with constraints on the number of false positive and false negative mistakes. We then use this online learner and a conversion technique from Helmbold et al. (2000a) to construct a randomized online learner in the realizable, *apple tasting* feedback setting with the stated guarantee in (2). For the lower bound in (2), we consider a new combinatorial object called an *apple tree* and use it to explicitly construct a hard, realizable stream for any randomized, apple tasting learner. This is in contrast to Helmbold et al. (2000a), who prove lower bounds on the minimax expected number of mistakes by converting randomized apple tasting learners into deterministic full-information feedback learners.

1.1. Related Works

Apple tasting is usually presented as an example of a more general partial feedback setting called *partial monitoring* games, where the player's feedback is specified by a feedback matrix (Cesa-Bianchi and Lugosi, 2006; Bartók et al., 2014). Of particular interest is the work by Bartók (2012), who gives a beautiful result (Theorem 2) characterizing the minimax rates in different partial monitoring games (including apple tasting). However, this is done for a slightly different setting where there is no hypothesis class $\mathcal{H}$, but just a *finite* set of actions the learner can play. The goal here is to compete with the best fixed *action* in hindsight. In contrast, in our setting, there is a hypothesis class, often *infinite* in size, and the goal is compete against the best fixed *hypothesis* in hindsight. Related to partial monitoring games is sequential prediction with *graph feedback*, for which apple tasting feedback is also special case (Alon et al., 2015). In this model, a learner plays a repeated game against an adversary. In each round, the learner selects one of K actions but observes the losses for a subset of the actions determined by a combinatorial structure called a *feedback graph*. Alon et al. (2015) classify feedback graphs into three types and establish a trichotomy on the rates of the minimax regret based on the type of graph. In this paper, we extend the online learner presented in Alon et al. (2015) to the setting of binary prediction with expert advice to establish the minimax regret of apple tasting in the agnostic setting.

In a parallel direction, there has been an explosion of work using combinatorial dimension to give tight quantitative characterizations of online learnability. For example, Littlestone (1987) proposed the Littlestone dimension and showed that it exactly characterizes the optimal mistake bound of deterministic learners for online binary classification in the full-information, realizable setting. Later, Ben-David et al. (2009) show that the Littlestone dimension also provides a tight quantitative characterization of the optimal expected regret in the full-information, agnostic setting. Later, Daniely et al. (2011) define a *multiclass* extension of the Littlestone dimension and show that it provides a tight quantitative characterization of realizable and agnostic mutliclass online learnability under full-information feedback when the label space is finite. In their same work, Daniely et al. (2011) define the Bandit Littlestone dimension and show that it exactly characterizes the optimal mistake bound of deterministic learners in the realizable setting under partial feedback setting known as bandit feedback. Daniely and Helbertal (2013) and Raman et al. (2024) later show that the Bandit Littlestone dimension also characterizes agnostic bandit online learnability. Beyond binary and multiclass classification, combinatorial dimensions have been used to characterize online learnability for regression (Rakhlin et al., 2015), list classification (Moran et al., 2023), ranking (Raman et al., 2023b), and general supervised online learning models (Raman et al., 2023a).

2. Preliminaries

2.1. Notation

Let $\mathcal{X}$ denote the instance space and $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ denote a binary hypothesis class. Given an instance $x \in \mathcal{X}$, and any collection of hypothesis $V \subseteq \{0,1\}^{\mathcal{X}}$, we let $V(x) := \{h(x) : h \in V\}$ denote the projection of V onto x. As usual, $[N]$ is used to denote $\{1, 2, \ldots, N\}$.

2.2. Online Classification and Apple Tasting

In the standard binary online classification setting with full-information feedback a learner $\mathcal{A}$ plays a repeated game against an adversary over T rounds. In each round $t \in [T]$, the adversary picks a labeled instance $(x_t, y_t) \in \mathcal{X} \times \{0,1\}$ and reveals x_t to the learner. The learner makes a (possibly randomized) prediction $\mathcal{A}(x_t) \in \{0,1\}$. Finally, the adversary reveals the true label y_t and the learner suffers the 0-1 loss $\mathbb{1}\{\mathcal{A}(x_t) \neq y_t\}$. Given a hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, the goal of the learner is to output predictions such that its *expected regret*

$$\mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) := \sup_{(x_1,y_1),\ldots,(x_T,y_T)} \left(\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{\mathcal{A}(x_t) \neq y_t\}\right] - \inf_{h \in \mathcal{H}} \sum_{t=1}^{T} \mathbb{1}\{h(x_t) \neq y_t\} \right)$$

is small, where the expectation is only over the randomness of the learner. A hypothesis class $\mathcal{H}$ is said to be online learnable under full-information feedback, if there exists an (potentially randomized) online learning algorithm $\mathcal{A}$ such that $\mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) = o(T)$ while $\mathcal{A}$ receives the true label y_t at the end of each round. If it is guaranteed that the learner always observes a sequence of examples labeled by some hypothesis $h \in \mathcal{H}$, then we say we are in the *realizable* setting and the goal of the learner is to minimize its *expected cumulative mistakes*,

$$\mathrm{M}_{\mathcal{A}}(T, \mathcal{H}) := \sup_{h \in \mathcal{H}} \sup_{x_1,\ldots,x_T} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{\mathcal{A}(x_t) \neq h(x_t)\}\right],$$

where again the expectation is taken only with respect to the randomness of the learner. In the apple tasting feedback model, the adversary still picks a labeled instance $(x_t, y_t) \in \mathcal{X} \times \{0,1\}$ and reveals x_t to the learner. However, the learner only gets to observe the true label y_t if they predict $\hat{y}_t = 1$. Analogous to the full-information setting, a hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ is online learnable under apple tasting feedback, if there exists an online learning algorithm whose expected regret, *under apple tasting feedback*, on any sequence of labeled instances is $o(T)$.

Definition 1 (Agnostic Online Learnability under Apple Tasting Feedback) *A hypothesis class $\mathcal{H}$ is online learnable under apple tasting feedback, if there exists an algorithm $\mathcal{A}$ such that $\mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) = o(T)$ while $\mathcal{A}$ only receives feedback when predicting* 1.

As in the full-information setting, if it is guaranteed that the sequence of examples is labeled by some hypothesis $h \in \mathcal{H}$, then we say we are in the *realizable* setting and an analogous definition of learnability under apple tasting feedback follows.

Definition 2 (Realizable Online Learnability under Apple Tasting Feedback) *A hypothesis class $\mathcal{H}$ is online learnable under apple tasting feedback in the* realizable setting*, if there exists an algorithm $\mathcal{A}$ such that $\mathrm{M}_{\mathcal{A}}(T, \mathcal{H}) = o(T)$ while $\mathcal{A}$ only receives feedback when predicting* 1.

2.3. Trees and Combinatorial Dimensions

In online learning, combinatorial dimensions are defined in terms of *trees*, a basic unit that captures temporal dependence. A binary tree $\mathcal{T}$ of depth d is *complete* if it admits the following recursive structure. A depth one complete binary tree is a single root node with left and right outgoing edges. A complete binary tree $\mathcal{T}$ of depth d has a root node whose left and right subtrees are each complete binary trees of depth $d-1$. Given a complete binary tree $\mathcal{T}$, we can label its internal nodes and edges by elements of $\mathcal{X}$ and $\{0,1\}$ respectively to get a *Littlestone tree*.

Definition 3 (Littlestone tree) *A Littlestone tree of depth d is a* complete *binary tree of depth d where the internal nodes are labeled by instances of $\mathcal{X}$ and the left and right outgoing edges from each internal node are labeled by 0 and 1 respectively.*

Given a Littlestone tree $\mathcal{T}$ of depth d, a root-to-leaf path down $\mathcal{T}$ is a bitstring $\sigma \in \{0,1\}^d$ indicating whether to go left ($\sigma_i = 0$) or to go right ($\sigma_i = 1$) at each depth $i \in [d]$. A path $\sigma \in \{0,1\}^d$ down $\mathcal{T}$ gives a sequence of labeled instances $\{(x_i, \sigma_i)\}_{i=1}^d$, where x_i is the instance labeling the internal node following the prefix $(\sigma_1, ..., \sigma_{i-1})$ down the tree. A hypothesis $h_\sigma \in \mathcal{H}$ shatters a path $\sigma \in \{0,1\}^d$, if for every $i \in [d]$, we have $h_\sigma(x_i) = \sigma_i$. In other words, h_σ is consistent with the labeled examples when following σ. A Littlestone tree $\mathcal{T}$ is shattered by $\mathcal{H}$ if for every root-to-leaf path σ down $\mathcal{T}$, there exists a hypothesis $h_\sigma \in \mathcal{H}$ that shatters it. Using this notion of shattering, we define the Littlestone dimension of a hypothesis class.

Definition 4 (Littlestone dimension) *The Littlestone dimension of $\mathcal{H}$, denoted $\mathrm{L}(\mathcal{H})$, is the largest $d \in \mathbb{N}$ such that there exists a Littlestone tree $\mathcal{T}$ of depth d shattered by $\mathcal{H}$. If there exists shattered Littlestone trees $\mathcal{T}$ of arbitrary depth, then we say that $\mathrm{L}(\mathcal{H}) = \infty$.*

Remarkably, the Littlestone dimension gives a tight quantitative characterization of realizable learnability under full-information feedback. In particular, Littlestone (1987) gives a generic deterministic algorithm, termed the Standard Optimal Algorithm (SOA), and shows that it makes at most $\mathrm{L}(\mathcal{H})$ number of mistakes on any realizable stream. Moreover, they showed that for every deterministic learner, there exists a realizable stream that can force at least $\mathrm{L}(\mathcal{H})$ mistakes, proving that the Ldim exactly quantifies the mistake bound for deterministic realizable learnability under full-information feedback.

Under apple tasting feedback, one can use the lower and upper bounds derived by Helmbold et al. (2000a) to deduce that the Ldim also provides a *qualitative* characterization of realizable learnability. However, unlike the full-information feedback setting, the Ldim alone cannot provide matching lower and upper bounds on the minimax expected number of mistakes under apple tasting feedback. Indeed, for the simple class of singletons over the natural numbers, $\mathcal{H}_{\text{sing}} := \{x \mapsto \mathbb{1}\{x = a\} : a \in \mathbb{N}\}$ we have that $\mathrm{L}(\mathcal{H}_{\text{sing}}) = 1$ while the minimax expected number of mistakes scales with the time horizon T (see Section 3.2). On the other hand, for the "flip" of the singletons, $\mathcal{H} = \{x \mapsto \mathbb{1}\{x \neq a\} : a \in \mathbb{N}\}$, we also have that $\mathrm{L}(\mathcal{H}) = 1$, but $\mathcal{H}$ is trivially learnable in at most 1 mistake in the realizable setting. Accordingly, new ideas are needed to handle the asymmetric nature of apple tasting feedback.

As a first step, we go beyond the symmetric nature of complete binary trees and define a new asymmetric binary tree called an *apple tree*. In particular, a binary tree $\mathcal{T}$ of depth d and width w is an apple tree if it admits the following recursive structure. An apple tree of width $w \geq d$ is a

complete binary tree with depth d. An apple tree with width $w = 1$ and depth d is a degenerate binary tree of depth d where every internal node has only a left child. An apple tree $\mathcal{T}(w, d)$ of depth d and width $w < d$ has a root node v whose left subtree is an apple tree $\mathcal{T}(w, d-1)$ and whose right subtree is an apple tree $\mathcal{T}(w-1, d-1)$. At a high-level, the width of an apple tree w controls the number of ones any path starting from the root can have before the path ends. The depth d of an apple tree controls the maximum number of zeros along any path starting from the root. From this perspective, one can alternatively construct an apple tree of width w and depth d by starting with a complete binary tree of depth d and then trimming each path starting from the root node such that it ends once it contains w ones or until a leaf node has been reached.

Similar to Littlestone trees, we can label the internal nodes of an apple tree with instances in $\mathcal{X}$ and the edges with elements of $\{0, 1\}$. By doing so, we get an Apple Littlestone (AL) tree.

Definition 5 (Apple Littlestone tree) *An* Apple Littlestone tree *of width w and depth d is an* apple *tree of width w and depth d where the internal nodes are labeled by instances of $\mathcal{X}$ and the left and right outgoing edges from each internal node are labeled by* 0 *and* 1 *respectively.*

The notion of shattering for Littlestone trees extends exactly to AL trees. Formally, an AL tree $\mathcal{T}(w, d)$ of width w and depth d is shattered by $\mathcal{H}$, if for every path σ down the tree $\mathcal{T}$, there exists a hypothesis $h_\sigma \in \mathcal{H}$ consistent with $\{(x_i, \sigma_i)\}_{i=1}^{|\sigma|}$. Note that, unlike Littlestone trees, AL trees are *imbalanced*. In fact, for an AL tree $\mathcal{T}$ of width w and depth d, there can be at most w ones along any valid path σ down the tree before the path ends. Therefore, not all root-to-leaf paths are of the same length. Nevertheless, this notion of shattering is still well defined and naturally leads to a combinatorial dimension analogous to the Littlestone dimension.

Definition 6 (Apple Littlestone dimension) *The* Apple Littlestone dimension *of $\mathcal{H}$ at width $w \in \mathbb{N}$, denoted $\mathrm{AL}_w(\mathcal{H})$, is the largest d such that there exists an apple tree $\mathcal{T}(w, d)$ of width w and depth d shattered by $\mathcal{H}$. If there exists shattered Apple Littlestone trees $\mathcal{T}$ with width w of arbitrarily large depth, then we say that $\mathrm{AL}_w(\mathcal{H}) = \infty$. If there are no shattered apple trees $\mathcal{T}$ of width w, then we say that $\mathrm{AL}_w(\mathcal{H}) = 0$.*

In general, the value of $\mathrm{AL}_w(\mathcal{H})$ for $w \leq \mathrm{L}(\mathcal{H})$ can be much larger than $\mathrm{L}(\mathcal{H})$. For example, even for the class of singletons defined over $\mathbb{N}$, we have that $\mathrm{AL}_1(\mathcal{H}_{\text{sing}}) = \infty$ while $\mathrm{L}(\mathcal{H}_{\text{sing}}) = 1$. Accordingly, unlike the Ldim, the Apple Littlestone dimension (ALdim), does not provide a qualitative characterization of learnability. Instead, using the ALdim, we define a new combinatorial parameter termed the Effective width. In Section 3, we show that the Effective width provides a tight *quantitative characterization* of realizable learnability under apple tasting feedback.

Definition 7 (Effective width) *The* Effective width *of a hypothesis class $\mathcal{H}$, denoted $\mathrm{W}(\mathcal{H})$, is the smallest $w \in \mathbb{N}$ such that $\mathrm{AL}_w(\mathcal{H}) < \infty$. If there is no $w \in \mathbb{N}$ such that $\mathrm{AL}_w(\mathcal{H}) < \infty$, then we say that $\mathrm{W}(\mathcal{H}) = \infty$.*

The following lemma, whose proof is in Appendix B, establishes important properties of $\mathrm{AL}_w(\mathcal{H})$ and $\mathrm{W}(\mathcal{H})$ that we use to characterize learnability.

Lemma 8 (Structural Properties) *For every $\mathcal{H} \subseteq \{0, 1\}^{\mathcal{X}}$, the following statements are true.*

(i) $\mathrm{AL}_{w_1}(\mathcal{H}) \geq \mathrm{AL}_{w_2}(\mathcal{H})$ *for all* $w_1 < w_2$.

(ii) $\mathrm{AL}_w(\mathcal{H}) \geq \min\{w, \mathrm{L}(\mathcal{H})\}$.

(iii) $\mathrm{AL}_w(\mathcal{H}) = \mathrm{L}(\mathcal{H})$ *for all* $w \geq \mathrm{L}(\mathcal{H}) + 1$ *when* $\mathrm{L}(\mathcal{H}) < \infty$.

(iv) $\mathrm{W}(\mathcal{H}) \leq \mathrm{L}(\mathcal{H}) + 1$ *when* $\mathrm{L}(\mathcal{H}) < \infty$.

(v) $\mathrm{W}(\mathcal{H}) < \infty \iff \mathrm{L}(\mathcal{H}) < \infty$.

Property (iv) can be tight in the sense that for the class of singletons, $\mathrm{W}(\mathcal{H}_{\text{sing}}) = 2$ while $\mathrm{L}(\mathcal{H}_{\text{sing}}) = 1$. Moreover, one cannot hope to lower bound $\mathrm{W}(\mathcal{H})$ in terms of $\mathrm{L}(\mathcal{H})$. Indeed, for any *finite* hypothesis class $\mathcal{H}$, we have that $\mathrm{W}(\mathcal{H}) = 1$ while $\mathrm{L}(\mathcal{H})$ can be made arbitrarily large. Finally, as an example, we also compute the Effective width for the k-wise generalization of $\mathcal{H}_{\text{sing}}$ in Appendix G.

3. Realizable Learnability

In this section, we revisit the learnability of apple tasting in the realizable setting, first studied by Helmbold et al. (2000a). Our main result is Theorem 9, which lower- and upper bounds the minimax expected number of mistakes in terms of the Littlestone dimension and the Effective width.

Theorem 9 (Realizable Learnability) *For any hypothesis class* $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$,

$$\frac{1}{8}\min\left\{\max\left\{\sqrt{(\mathrm{W}(\mathcal{H})-1)\,T}, \mathrm{L}(\mathcal{H})\right\}, T\right\} \leq \inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T,\mathcal{H}) \leq \mathrm{AL}_{\mathrm{W}(\mathcal{H})}(\mathcal{H}) + 2\sqrt{(\mathrm{W}(\mathcal{H})-1)T}.$$

The lower and upper bounds of Theorem 9 can be tight up to constant factors. There are two cases to consider. When $\mathrm{W}(\mathcal{H}) = 1$, the lower and upper bounds in Theorem 9 reduce to $\frac{\mathrm{L}(\mathcal{H})}{8} \leq \inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T,\mathcal{H}) \leq \mathrm{AL}_1(\mathcal{H})$ for $T \geq \mathrm{L}(\mathcal{H})$. Taking $|\mathcal{X}| = d < \infty$ and $\mathcal{H} = \{0,1\}^{\mathcal{X}}$ gives that $\mathrm{L}(\mathcal{H}) = \mathrm{AL}_1(\mathcal{H}) = d$, ultimately implying that the lower- and upper bounds can be off by only a constant factor of $\frac{1}{8}$. Secondly, consider the case where $\mathrm{W}(\mathcal{H}) \geq 2$. Then, if $T \geq \max\{\mathrm{W}(\mathcal{H}) - 1, \mathrm{AL}^2_{\mathrm{W}(\mathcal{H})}(\mathcal{H})\}$, Theorem 9 implies that $\frac{1}{8}\sqrt{(\mathrm{W}(\mathcal{H})-1)\,T} \leq \inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T,\mathcal{H}) \leq 3\sqrt{(\mathrm{W}(\mathcal{H})-1)T}$, showing that the upper- and lower bounds are off only by a constant factor.

Theorem 9 implies that when $\mathrm{W}(\mathcal{H}) = 1$, a constant upper bound on the expected regret is possible. In fact, when $\mathrm{AL}_1(\mathcal{H}) < \infty$, there exists a *deterministic* online learner which makes at most $\mathrm{AL}_1(\mathcal{H})$ mistakes in the realizable setting under apple tasting feedback (see Appendix A). On the other hand, Theorem 9 also shows that, in full generality, it is not possible to achieve a constant *expected* mistake bound under apple tasting feedback in the *realizable* setting. Indeed, if $\mathrm{W}(\mathcal{H}) > 1$, then the worst-case expected mistakes of any randomized learner, under apple tasting feedback, is at least $\Omega(\sqrt{T})$. This is in contrast to the full-information setting, where the minimax expected number of mistakes in the realizable setting is constant, and that too achieved by a *deterministic* learner (i.e SOA). Accordingly, Theorem 9 gives a trichotomy in the minimax expected number of mistakes for the realizable setting.

Corollary 10 (Trichotomy in minimax expected number of mistakes) *For any hypothesis class* $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$,

$$\inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T,\mathcal{H}) = \begin{cases} \Theta(1), & \text{if } \mathrm{W}(\mathcal{H}) = 1. \\ \Theta(\sqrt{T}) & \text{if } 2 \leq \mathrm{W}(\mathcal{H}) < \infty. \\ \Theta(T), & \mathrm{W}(\mathcal{H}) = \infty. \end{cases}$$

In Section 4, we will show that $\inf_{\mathcal{A}} \mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) = \tilde{\Theta}(\sqrt{T})$, where $\tilde{\Theta}$ hides poly-logarithmic factors in T. With this in mind, Corollary 10 shows that when $\mathrm{W}(\mathcal{H}) \geq 2$, realizable learnability under apple tasting feedback can be as hard as agnostic learnability under apple tasting feedback. Unfortunately, for many simple classes, like the singletons over $\mathbb{N}$, we have $\mathrm{W}(\mathcal{H}) \geq 2$. On the other hand, for classes containing hypothesis that rarely output 0, like the "flip" of the class of singletons, realizable learnability under apple tasting feedback can be as easy as realizable learnability under full-information feedback.

3.1. Upper Bounds for Randomized Learners in the Realizable Setting

We prove a slightly stronger upper bound than the one stated in Theorem 9.

Lemma 11 (Randomized Realizable Upper Bound) *For any hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$,*

$$\inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T, \mathcal{H}) \leq \inf_{w \in \mathbb{N}} \left\{ \mathrm{AL}_w(\mathcal{H}) + 2\sqrt{(w-1)T} \right\}.$$

The upper bound in Theorem 9 follows by picking $w = \mathrm{W}(\mathcal{H})$. If one picks $w = \mathrm{L}(\mathcal{H}) + 1$, then $\mathrm{AL}_w(\mathcal{H}) = \mathrm{L}(\mathcal{H})$ and we get an upper bound of $3\sqrt{\mathrm{L}(\mathcal{H})T}$ on the expected mistakes.

Lemma 11 follows from composing the next two lemmas. Lemma 12 shows that if $\mathrm{AL}_w(\mathcal{H}) < \infty$, then there exists a deterministic online learner, under *full-information* feedback, that makes at most $w-1$ false negative mistakes and at most $\mathrm{AL}_w(\mathcal{H})$ false positive mistakes. Lemma 13 is from Helmbold et al. (2000a) and shows how to convert any online learner under full-information feedback into an online learner under apple tasting feedback.

Lemma 12 *For any $\mathcal{H} \subset \{0,1\}^{\mathcal{X}}$ and $w \in \mathbb{N}$ such that $\mathrm{AL}_w(\mathcal{H}) < \infty$, there exists a deterministic online learner which, under* full-information *feedback, makes at most $w-1$* false negative *mistakes and at most $\mathrm{AL}_w(\mathcal{H})$* false positive *mistakes in the realizable setting.*

Proof Suppose $w \in \mathbb{N}$ such that $\mathrm{AL}_w(\mathcal{H}) < \infty$ and denote $\mathcal{A}$ to be Algorithm 1.

Algorithm 1 Realizable Algorithm Under Full-Information Feedback

Input: $V_1 = \mathcal{H}$, pick $w_1 = w$ such that $\mathrm{AL}_w(\mathcal{H}) < \infty$
for $t = 1, ..., T$ **do**
  Receive x_t.
  For each $y \in \{0,1\}$, define $V_t^y = \{h \in V_t \mid h(x_t) = y\}$.
  if $V_t(x_t) = \{y\}$ **then**
    Predict $\hat{y}_t = y$.
  else
    If $|V_t^1| \geq 1$, and $\mathrm{AL}_{w_t}(V_t^0) < \mathrm{AL}_{w_t}(V_t)$, predict $\hat{y}_t = 1$. Otherwise, predict $\hat{y}_t = 0$.
  Receive y_t and update $V_t \leftarrow V_t^{y_t}$.
  If $\hat{y}_t = 0$ and $y_t = 1$, then update $w_{t+1} \leftarrow w_t - 1$. Else, set $w_{t+1} \leftarrow w_t$.
end

Let $(x_1, y_1), ..., (x_T, y_T)$ be the stream to be observed by $\mathcal{A}$. We show that $\mathcal{A}$, initialized at $w_1 = w$, makes at most $\mathrm{AL}_w(\mathcal{H})$ false positive mistakes and at most $w-1$ false negative mistakes. Let $S_+ = \{t \in [T] \mid \hat{y}_t = 1 \text{ and } y_t = 0\}$ be the set of time points where $\mathcal{A}$ makes false positive

mistakes, and $S_- = \{t \in [T] \mid \hat{y}_t = 0 \text{ and } y_t = 1\}$ be the set of time points where $\mathcal{A}$ makes false negative mistakes. We show $|S_+| \leq \mathrm{AL}_w(\mathcal{H})$ by first establishing

$$\mathrm{AL}_{w_{t+1}}(V_{t+1}) \leq \mathrm{AL}_{w_t}(V_t) - \mathbb{1}\{t \in S_+\}, \qquad \forall t \in [T]. \tag{1}$$

This inequality then implies that the number of false positive mistakes of $\mathcal{A}$ is

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{1}\{t \in S_+\} &\leq \sum_{t=1}^{T} \left(\mathrm{AL}_{w_t}(V_t) - \mathrm{AL}_{w_{t+1}}(V_{t+1})\right) \\ &= \mathrm{AL}_{w_1}(V_1) - \mathrm{AL}_{w_{T+1}}(V_{T+1}) \\ &\leq \mathrm{AL}_{w_1}(V_1) = \mathrm{AL}_w(\mathcal{H}).\end{aligned}$$

To prove inequality (1), we consider the two cases: $t \in S_+$ and $t \notin S_+$. Suppose $t \in S_+$. Then, we know that $\hat{y}_t = 1$ and by the prediction rule of $\mathcal{A}$, we must have $\mathrm{AL}_{w_t}(V_t^0) < \mathrm{AL}_{w_t}(V_t)$. Since $y_t = 0$, we further obtain that $V_{t+1} = V_t^0$ and $w_{t+1} = w_t$ in this case. This yields $\mathrm{AL}_{w_{t+1}}(V_{t+1}) < \mathrm{AL}_{w_t}(V_t)$, which subsequently implies $\mathrm{AL}_{w_{t+1}}(V_{t+1}) \leq \mathrm{AL}_{w_t}(V_t) - \mathbb{1}\{t \in S_+\}$.

Now, let us consider the case when $t \notin S_+$. In the case when $t \notin S_+ \cup S_-$, we have $w_{t+1} = w_t$ and $\mathbb{1}\{t \in S_+\} = 0$. Thus, we trivially obtain $\mathrm{AL}_{w_{t+1}}(V_{t+1}) \leq \mathrm{AL}_{w_t}(V_t) - \mathbb{1}\{t \in S_+\}$ since $V_{t+1} \subseteq V_t$. Next, let us consider the case when $t \in S_-$. In this case, we have $w_{t+1} = w_t - 1$, $V_t = V_t^1$, and $\mathbb{1}\{t \in S_+\} = 0$. Thus, to establish inequality (1), it suffices to show that $\mathrm{AL}_{w_t-1}(V_t^1) \leq \mathrm{AL}_{w_t}(V_t)$. Suppose, for the sake of contradiction, this is not true and we instead have $\mathrm{AL}_{w_t-1}(V_t^1) > \mathrm{AL}_{w_t}(V_t)$. Let $d := \mathrm{AL}_{w_t}(V_t)$. Note that $d > 0$ because there must exist $h_1, h_2 \in V_t$ such that $h_1(x_t) \neq h_2(x_t)$ or otherwise $\mathcal{A}$ would not have made a false negative mistake. Since $\mathrm{AL}_{w_t-1}(V_t^1) > d$, we are guaranteed the existence of an AL tree $\mathcal{T}_1(w_t - 1, d)$ shattered by V_t^1. Furthermore, as $\hat{y}_t = 0$ and $|V_t^1| \geq 1$, the prediction rule implies that $\mathrm{AL}_{w_t}(V_t^0) \geq \mathrm{AL}_{w_t}(V_t) = d$. Accordingly, we are also guaranteed the existence of an AL tree $\mathcal{T}_0(w_t, d)$ shattered by V_t^0. Now consider an AL tree $\mathcal{T}$ that has x_t in its root-node, has a subtree $\mathcal{T}_0(w_t, d)$ attached to left-outgoing edge from the root-node and has a subtree $\mathcal{T}_1(w_t - 1, d)$ attached to right-outgoing edge from the root-node. Since all hypotheses in V_t^0 output 0 on x_t and all hypotheses in V_t^1 output 1 on x_t, the tree $\mathcal{T}$ shattered by V_t. Since $\mathcal{T}$ is a valid AL tree of width w_t and depth $d + 1$, we have that $\mathrm{AL}_{w_t}(V_t) \geq d + 1$, a contradiction to our assumption that $\mathrm{AL}_{w_t}(V_t) = d$. Therefore, we must have $\mathrm{AL}_{w_t-1}(V_t^1) \leq \mathrm{AL}_{w_t}(V_t)$ when $t \in S_-$.

Next, we show that $\mathcal{A}$ makes at most $w - 1$ false negative mistakes. Let $t^\star \in [T]$ be the time point where the algorithm makes its $(w - 1)$-th false negative mistake. If such time point $t^\star$ does not exist, then we trivially have $|S_-| \leq w - 2 < w - 1$. We now consider the case when $t^\star \in [T]$ exists. It suffices to show that, $\forall t > t^\star$, we have $t \notin S_-$. Suppose, for the sake of contradiction, $\exists\, t > t^\star$ such that $t \in S_-$. Since $\hat{y}_t = 0$ and $y_t = 1$, we must have $|V_t^1| \geq 1$. Thus, the prediction strategy implies that $\mathrm{AL}_{w_t}(V_t^0) \geq \mathrm{AL}_{w_t}(V_t)$. Given that $t > t^\star$ and $\mathcal{A}$ has already made $w - 1$ false negative mistakes, we must have $w_t = 1$. Thus, we have $\mathrm{AL}_1(V_t^0) \geq \mathrm{AL}_1(V_t) =: d$. Note that $d \geq 1$ because there must exist $h_1, h_2 \in V_t$ such that $h_1(x_t) \neq h_2(x_t)$. Since $\mathrm{AL}_1(V_t^0) \geq d$, we are guaranteed the existence of an AL tree $\mathcal{T}_0(1, d)$ of width 1 and depth d shattered by V_t^0. Next, consider a tree $\mathcal{T}$ with x_t on the root node and has a subtree $\mathcal{T}_0(1, d)$ attached to the left-outgoing edge from the root node. Let $h \in V_t$ any hypothesis such that $h(x_t) = 1$. The hypothesis h must exist because $|V_t^1| \geq 1$. By putting h in the leaf node following the right-outgoing edge from the root node in $\mathcal{T}$, it is clear that $\mathcal{T}$ is a valid AL tree of width 1 and depth $d + 1$ shattered by V_t.

The existence of $\mathcal{T}$ implies that $\mathrm{AL}_1(V_t) \geq d+1$, a contradiction to our assumption $\mathrm{AL}_1(V_t) = d$. Thus, $\forall t > t^\star$, we have $t \notin S_-$. Therefore, $\mathcal{A}$ makes no more than $w-1$ false negative mistakes. ■

We remark that Helmbold et al. (2000b) also give a deterministic online learner in the full-information setting under constraints on the number of false positive and false negative mistakes (see Algorithm SCS in (Helmbold et al., 2000b, Section 2)). However, similar to Helmbold et al. (2000a), their algorithm checks the existence of an online learning algorithm satisfying certain properties. We extend on this result by giving an SOA-type algorithm that only requires computing combinatorial dimensions.

Lemma 13 is the restatement of Corollary 2 in Helmbold et al. (2000a). For completeness sake, we provide a proof in Appendix C. Lemma 11 follows by composing Lemma 12 and Lemma 13.

Lemma 13 (Helmbold et al. (2000a)) *For any $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, if there exists a deterministic learner which, under full-information feedback, makes at most M_-* false negative *mistakes and at most M_+* false positive *mistakes, then there exists a randomized learner, whose expected number of mistakes, under apple tasting feedback, is at most $M_+ + 2\sqrt{TM_-}$ in the realizable setting.*

3.2. Lower Bounds for Randomized Learners in the Realizable Setting

As in the upper bound, we prove a slightly stronger lower bound than the one stated in Theorem 9.

Lemma 14 (Realizable Lower Bound) *For any hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$,*

$$\inf_{\mathcal{A}} \mathrm{M}_{\mathcal{A}}(T, \mathcal{H}) \geq \frac{1}{8} \sup_{w \in \mathbb{N}} \sqrt{\min\{w, \mathrm{L}(\mathcal{H}), T\} \, \min\{\mathrm{AL}_w(\mathcal{H}), T\}}.$$

The lower bounds in Theorem 9 follows by picking $w = \mathrm{W}(\mathcal{H}) - 1$ and $w = \mathrm{L}(\mathcal{H}) + 1$ respectively. When $w = \mathrm{W}(\mathcal{H}) - 1$, we have that $\min\{w, \mathrm{L}(\mathcal{H}), T\} = \min\{\mathrm{W}(\mathcal{H}) - 1, T\}$ and $\min\{\mathrm{AL}_w(\mathcal{H}), T\} \geq \min\{\mathrm{W}(\mathcal{H}) - 1, T\}$ using Lemma 8 (ii) and (iv). On the other hand, when $w = \mathrm{L}(\mathcal{H}) + 1$, we have that $\min\{\mathrm{AL}_w(\mathcal{H}), T\} = \min\{\mathrm{L}(\mathcal{H}), T\}$ using Lemma 8 (iii).

Proof Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, $w \in \mathbb{N}$, and $T \in \mathbb{N}$ be the time horizon. Since learning under apple tasting feedback implies learning under full-information feedback, a lower bound of $\frac{\min\{T, \mathrm{L}(\mathcal{H})\}}{2}$ on the minimax expected number of mistakes follows trivially from the full-information feedback lower bound. Accordingly, for the remainder of the proof we suppose $w \leq \min\{\mathrm{L}(\mathcal{H}), T\}$, since if this condition is not met, the claimed lower bound is at most $\frac{\min\{T, \mathrm{L}(\mathcal{H})\}}{2}$. Let $\mathcal{T}$ be any AL tree of width w of depth $d = \left\lfloor \sqrt{w \min\{T, \mathrm{AL}_w(\mathcal{H})\}} \right\rfloor$ shattered by $\mathcal{H}$. Such a tree must exist because $d \leq \mathrm{AL}_w(\mathcal{H})$. Let $\mathcal{A}$ be any randomized apple tasting online learner. Our goal will be to construct a hard, determinsitic, realizable stream of instances $(x_1, y_1), ..., (x_T, y_T)$ such that $\mathcal{A}$'s expected regret is at least $\frac{d}{4}$.

We first construct a path $\sigma^\star$ down $\mathcal{T}$ recursively using $\mathcal{A}$. Starting with $\sigma^\star_1$, let A_1 be the event that $\mathcal{A}$, if presented with $\lfloor \frac{d}{w} \rfloor$ copies of the root node $x^\star_1$, predicts 1 on at least one of the copies. Then, set $\sigma^\star_1 = 0$ if $\mathbb{P}(A_1) \geq \frac{1}{2}$ and set $\sigma^\star_1 = 1$ otherwise. For $j \geq 2$, let $x^\star_1, ..., x^\star_j$ be the sequence of instances labeling the internal nodes along the prefix $(\sigma^\star_1, ..., \sigma^\star_{j-1})$ down $\mathcal{T}$. Let A_j be the event that $\mathcal{A}$, if simulated with the sequence of $(j-1)\lfloor \frac{d}{w} \rfloor$ labeled instances consisting of $\lfloor \frac{d}{w} \rfloor$ copies of the labeled instance $(x^\star_1, \sigma^\star_1)$, followed by $\lfloor \frac{d}{w} \rfloor$ copies of the labeled instance $(x^\star_2, \sigma^\star_2)$,..., followed by $\lfloor \frac{d}{w} \rfloor$ copies of the labeled instance $(x^\star_{j-1}, \sigma^\star_{j-1})$, predicts the label 1 at least once when presented

with $\lfloor \frac{d}{w} \rfloor$ copies of the instance $x_j^\star$. Set $\sigma_j^\star = 0$ if $\mathbb{P}(A_j) \geq \frac{1}{2}$ and set $\sigma_j^\star = 1$ otherwise. Continue this process until $\sigma^\star$ is a valid path that reaches the end of tree $\mathcal{T}$.

We now construct our hard, labeled stream in blocks of size $\lfloor \frac{d}{w} \rfloor$. Each block only contains a single labeled instance, repeated $\lfloor \frac{d}{w} \rfloor$ times. For the first block B_1, repeat the labeled instance $(x_1^\star, \sigma_1^\star)$. Likewise, for block B_j for $2 \leq j \leq |\sigma^\star|$, repeat for $\lfloor \frac{d}{w} \rfloor$ times the labeled instance $(x_j^\star, \sigma_j^\star)$. Now, consider the stream $S = (B_1, ..., B_{|\sigma^\star|})$ obtained by concatenating the blocks $B_1, ..., B_{|\sigma^\star|}$ in that order. If $|\sigma^\star| \lfloor \frac{d}{w} \rfloor < T$, populate the rest of the stream S with the labeled instance $(x_{|\sigma^\star|}^\star, \sigma_{|\sigma^\star|}^\star)$.

We first claim that such a stream is realizable by $\mathcal{H}$. This follows trivially from the fact that (1) $\sigma^\star$ is a valid path down $\mathcal{T}$, (2) by the definition of shattering, there exists a hypothesis $h \in \mathcal{H}$ such that for all $j \in [|\sigma^\star|]$, we have $h(x_j^\star) = \sigma_j^\star$ and (3) our stream S only contains labeled instances from the set $\{(x_j^\star, \sigma_j^\star)\}_j$. We now claim that $\mathcal{A}$'s expected regret on the stream S is at least $\frac{d}{4}$. To see this, observe that whenever $\sigma_j^\star = 1$, $\mathcal{A}$'s expected mistakes on the block B_j is at least $\frac{1}{2} \lfloor \frac{d}{w} \rfloor$ since $\mathcal{A}$ gets passed the labeled instance $(x_j^\star, 1)$ for $\lfloor \frac{d}{w} \rfloor$ iterations, but the probability that it never predicts 1 on this batch after seeing $B_1, ..., B_{j-1}$ is $\mathbb{P}(A_j^c) \geq \frac{1}{2}$. Likewise, whenever $\sigma_j^\star = 0$, $\mathcal{A}$'s expected mistakes on the block B_j is at least $\frac{1}{2}$ since it gets passed the labeled instance $(x_j^\star, 0)$ for $\lfloor \frac{d}{w} \rfloor$ time points but predicts 1 on at least one of them with probability $\mathbb{P}(A_j) \geq \frac{1}{2}$.

We now lower bound the expected mistakes of $\mathcal{A}$ on the entire stream S by considering the number of ones in $\sigma^\star$ on a case by case basis. Note that since $\sigma^\star$ is a valid path down $\mathcal{T}$, we have $w \leq |\sigma^\star| \leq d$. Consider the case where $\sigma^\star$ has w ones. Then, $\mathcal{A}$'s expected regret is at least its expected regret on those batches B_j where $\sigma_j^\star = 1$. Thus, its expected regret is at least $\frac{w}{2} \lfloor \frac{d}{w} \rfloor \geq \frac{w}{2} \frac{d}{2w} \geq \frac{d}{4}$. Consider the case where $\sigma^\star$ has $w - j$ ones for $w \geq j \geq 1$. Then, since $\sigma^\star$ is a valid path, it must be the case that there are $d - (w - j)$ zero's in $\sigma^\star$. Therefore, $\mathcal{A}$'s expected regret is at least

$$\frac{(w-j)}{2} \left\lfloor \frac{d}{w} \right\rfloor + \frac{d - w + j}{2} \geq \frac{d}{2} - \frac{w-j}{2} + \frac{w-j}{2} \left\lfloor \frac{d}{w} \right\rfloor \geq \frac{d}{2}.$$

where the last inequality follows from the fact that $d \geq w$. Thus, in all cases, $\mathcal{A}$'s expected regret is at least $\frac{d}{4}$. The claimed lower bound follows by using the fact that $d \geq \sqrt{w \min\{T, \mathrm{AL}_w(\mathcal{H})\}}/2$. ■

4. Agnostic Learnability

We show that the Ldim quantifies the minimax expected regret in the agnostic setting under apple tasting feedback, closing the open problem posed by (Helmbold et al., 2000a, Page 138).

Theorem 15 (Agnostic Learnability) *For any hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$,*

$$\sqrt{\frac{\mathrm{L}(\mathcal{H})T}{8}} \leq \inf_{\mathcal{A}} \mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) \leq 3\sqrt{\mathrm{L}(\mathcal{H})T \ln T}.$$

The lower bound in Theorem 15 follows directly from the full-information lower bound in the agnostic setting (Ben-David et al., 2009). Therefore, in this section, we only focus on proving the upper bound. Our strategy will be in two steps. First, we modify the celebrated Randomized Exponential Weights Algorithm (REWA) (Cesa-Bianchi and Lugosi, 2006) to handle apple tasting

feedback by using the ideas from Alon et al. (2015). In particular, our algorithm EXP4.AT is an adaptation of EXP3.G from Alon et al. (2015) to binary prediction with expert advice under apple tasting feedback. Second, we give an agnostic online learner which uses the SOA to construct a finite set of experts that exactly covers $\mathcal{H}$ and then runs EXP4.AT using these experts. The upper bound in Theorem 15 follows immediately from the composition of these two results.

4.1. The EXP4.AT Algorithm

In this subsection, we present EXP4.AT, an adaptation of REWA to handle apple tasting feedback.

Algorithm 2 EXP4.AT: online learning with apple tasting feedback

Input: Learning rate $\eta \in (0, \frac{1}{2})$
Let q_1 be the uniform distribution over $[N]$
for $t = 1, ..., T$ **do**
  Get advice $\mathcal{E}_t^1, ..., \mathcal{E}_t^N \in \{0,1\}^N$
  Compute $p_t^1 = (1-\eta)\sum_{i=1}^N q_t^i \mathcal{E}_t^i + \eta$
  Predict $\hat{y}_t = 1$ with probability p_t^1 and $\hat{y}_t = 0$ with probability $p_t^0 = 1 - p_t^1$
  Observe true label y_t if $\hat{y}_t = 1$ and let $\hat{\ell}_t(y) = \frac{\mathbb{1}\{y \neq y_t\}\mathbb{1}\{\hat{y}_t = 1\}}{p_t^1}$
  For $i = 1, ..., N$ update $q_{t+1}^i = \frac{q_t^i \exp(-\eta \hat{\ell}_t(\mathcal{E}_t^i))}{\sum_{j=1}^N q_t^j \exp(-\eta \hat{\ell}_t(\mathcal{E}_t^j))}$
end

Theorem 16 (EXP4.AT Regret Bound) *If $\eta = \sqrt{\frac{\ln N}{2T}}$, then for any sequence of true labels $y_1, ..., y_T$, the predictions $\hat{y}_1, ..., \hat{y}_T$, output by* EXP4.AT *satisfy:*

$$\mathbb{E}\left[\sum_{t=1}^T \mathbb{1}\{\hat{y}_t \neq y_t\}\right] \leq \inf_{j \in [N]} \sum_{t=1}^T \mathbb{1}\{\mathcal{E}_t^j \neq y_t\} + 3\sqrt{T \ln N}.$$

In order to prove Theorem 16, we need the following lemma which gives a second-order regret bound for the EXP4.AT algorithm. The proof of Lemma 17 follows a similar potential-function strategy as in the proof of Lemma 4 in Alon et al. (2015) and can be found in Appendix D.

Lemma 17 (EXP4.AT Second-order Regret Bound) *For any $\eta \in (0, \frac{1}{2})$ and any sequence of true labels $y_1, ..., y_T$, the probabilities $p_1, ..., p_T$ output by* EXP4.AT *satisfy*

$$\sum_{t=1}^T \sum_{y \in \{0,1\}} p_t^y \hat{\ell}_t(y) - \inf_{j \in [N]} \sum_{t=1}^T \hat{\ell}_t(\mathcal{E}_t^j) \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \hat{\ell}_t(1) + \eta \sum_{t=1}^T p_t^1 (1 - p_t^1) \hat{\ell}_t(0)^2 + \eta \sum_{t=1}^T p_t^1 \hat{\ell}_t(1)^2.$$

Theorem 16 follows by taking expectations of both sides of the inequality in Lemma 17. The full proof can be found in Appendix E.

4.2. Proof Sketch of Theorem 15

Given any hypothesis class $\mathcal{H}$, we construct an agnostic online learner under apple tasting feedback with the claimed upper bound on expected regret. Similar to the generic agnostic online learner in the full-information setting (Ben-David et al., 2009), the high-level strategy is to use the SOA to construct a small set of experts E such that $|E| \leq T^{\mathrm{L}(\mathcal{H})}$ and for every $h \in \mathcal{H}$, there exists an expert $\mathcal{E}_h \in E$ such that $\mathcal{E}_h(x_t) = h(x_t)$ for all $t \in [T]$. Then, our agnostic online learner will run EXP4.AT using this set of experts E. The upper bound in Theorem 15 immediately follows from the guarantee of EXP4.AT in Theorem 16 and the fact that we have constructed an exact cover of $\mathcal{H}$. The full proof of Theorem 15 can be found in Appendix F.

5. Discussion and Open Questions

In this work, we revisited the classical setting of apple tasting and studied learnability from a combinatorial perspective. Our work makes an important step towards developing learning theory for online classification under partial feedback. An important future direction is to extend this work to multiclass classification under various partial feedback models, such as those captured by feedback graphs (Alon et al., 2015).

With respect to apple tasting, there are still interesting open questions. For example, our focus in the realizable setting was on *randomized* learnability. Remarkably, under full-information feedback, randomness is not needed to design online learners with optimal mistake bounds (up to constant factors). It is therefore natural to ask whether randomness is actually needed in the realizable setting under apple tasting feedback.

Question 1. For any $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ with $\mathrm{W}(\mathcal{H}) < \infty$, is $\inf_{\text{Deterministic } \mathcal{A}} \mathrm{M}_{\mathcal{A}}(T, \mathcal{H}) = o(T)$?

In Appendix A, we provide some partial answers. We show that if $\mathrm{W}(\mathcal{H}) = 1$ or $\mathrm{L}(\mathcal{H}) = 1$, then such generic deterministic learners do exist with mistake bounds that are constant factors away from the lower bound in Theorem 14. We conjecture that the statement in the open question is true.

Our lower and upper bounds in the agnostic setting are matching up to a factor logarithmic in T. Recently, Alon et al. (2021) showed that in the full-information setting, this $\log(T)$ factor can be removed from the upper bound, meaning that the optimal expected regret in the agnostic setting under full-information feedback is $\Theta(\sqrt{\mathrm{L}(\mathcal{H})T})$. As an open question, we ask whether it is possible to also remove the factor of $\log(T)$ from our upper bound in Theorem 15.

Question 2. For any $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, is it true that $\inf_{\mathcal{A}} \mathrm{R}_{\mathcal{A}}(T, \mathcal{H}) = \Theta(\sqrt{\mathrm{L}(\mathcal{H})T})$?

Acknowledgments

AT acknowledges the support of NSF via grant IIS-2007055. VR acknowledges the support of the NSF Graduate Research Fellowship. US acknowledges the support of the Rackham International Student Fellowship.

References

Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Conference on Learning Theory*, pages 23–35. PMLR, 2015.

Noga Alon, Omri Ben-Eliezer, Yuval Dagan, Shay Moran, Moni Naor, and Eylon Yogev. Adversarial laws of large numbers and optimal regret in online classification. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 447–455, 2021.

Gábor Bartók. The role of information in online learning. 2012.

Gábor Bartók, Dean P Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39 (4):967–997, 2014.

Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *COLT*, volume 3, page 1, 2009.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Amit Daniely and Tom Helbertal. The price of bandit information in multiclass online classification. In *Conference on Learning Theory*, pages 93–104. PMLR, 2013.

Amit Daniely, Sivan Sabato, Shai Ben-David, and Shai Shalev-Shwartz. Multiclass learnability and the erm principle. In Sham M. Kakade and Ulrike von Luxburg, editors, *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pages 207–232, Budapest, Hungary, 09–11 Jun 2011. PMLR.

David P Helmbold, Nicholas Littlestone, and Philip M Long. Apple tasting. *Information and Computation*, 161(2):85–139, 2000a.

David P Helmbold, Nicholas Littlestone, and Philip M Long. On-line learning with linear loss constraints. *Information and Computation*, 161(2):140–171, 2000b.

Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1987.

Shay Moran, Ohad Sharon, Iska Tsubari, and Sivan Yosebashvili. List online classification. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 1885–1913. PMLR, 2023.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *J. Mach. Learn. Res.*, 16(1):155–186, 2015.

Ananth Raman, Vinod Raman, Unique Subedi, Idan Mehalel, and Ambuj Tewari. Multiclass online learnability under bandit feedback. In *Proceedings of 35th International Conference on Algorithmic Learning Theory*, 2024. accepted.

Vinod Raman, Unique Subedi, and Ambuj Tewari. A combinatorial characterization of online learning games with bounded losses. *arXiv preprint arXiv:2307.03816*, 2023a.

Vinod Raman, Unique Subedi, and Ambuj Tewari. Online learning with set-valued feedback. *arXiv preprint arXiv:2306.06247*, 2023b.

Appendix A. Upper bounds for Deterministic Learners in the Realizable Setting

In this section, we provide deterministic apple tasting learners for some special classes. Our first contribution shows that when $\mathrm{W}(\mathcal{H}) = 1$, there exists deterministic online learner which makes at most $\mathrm{AL}_1(\mathcal{H})$ mistakes under apple tasting feedback.

Theorem 18 (Deterministic Realizable upper bound when $\mathrm{W}(\mathcal{H}) = 1$) *For any $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, there exists a deterministic online learner which, under apple tasting feedback, makes at most $\mathrm{AL}_1(\mathcal{H})$ mistakes in the realizable setting.*

Proof We will show that Algorithm 3 makes at most $\mathrm{AL}_1(\mathcal{H})$ mistakes in the realizable setting.

Algorithm 3 Deterministic Realizable Algorithm For Apple Tasting

Input: $V_1 = \mathcal{H}$
for $t = 1, ..., T$ **do**
  Receive x_t.
  If there exists $h \in V_t$ such that $h(x_t) = 1$, predict $\hat{y}_t = 1$. Else, predict $\hat{y}_t = 0$.
  If $\hat{y}_t = 1$, receive y_t and update $V_{t+1} \leftarrow \{h \in V_t : h(x_t) = y_t\}$
end

Let $t \in [T]$ be any round such that $\hat{y}_t \neq y_t$. We will show $\mathrm{AL}_1(V_{t+1}) \leq \mathrm{AL}_1(V_t) - 1$. By the prediction strategy and the fact that we are in the realizable setting, if $\hat{y}_t \neq y_t$ then it must be the case that $\hat{y}_t = 1$ but $y_t = 0$. For the sake of contradiction, suppose that $\mathrm{AL}_1(V_{t+1}) = \mathrm{AL}_1(V_t) = d$. Then, there exists an AL tree $\mathcal{T}$ of width 1 and depth d shattered by V_{t+1}. Consider a new AL tree $\mathcal{T}'$ of width 1 where the root node labeled is x_t and the left subtree of the root node is $\mathcal{T}$. Note that $\mathcal{T}'$ is a width 1 AL tree with depth $d + 1$. Since $\hat{y}_t = 1$, there exists a hypothesis $h \in V_t$ such that $h(x_t) = 1$. Moreover, for every hypothesis in $h \in V_{t+1} \subset V_t$, we have that $h(x_t) = 0$. Since $\mathcal{T}$ is shattered by $V_{t+1} \subset V_t$ and $\mathcal{T}$ is the left subtree of the root node in $\mathcal{T}'$, we have that $\mathcal{T}'$ is an AL tree of width 1 and depth $d + 1$ shattered by V_t. However, this contradicts our assumption that $\mathrm{AL}_1(V_t) = d$. Thus, it must be the case $\mathrm{AL}_1(V_{t+1}) \leq \mathrm{AL}_1(V_t) - 1$ whenever the algorithm errs, and the algorithm can err at most $\mathrm{AL}_1(\mathcal{H})$ times before $\mathrm{AL}_1(V_t) = 0$. ■

We extend the results of Theorem 18 to hypothesis classes where $\mathrm{L}(\mathcal{H}) = 1$. Note that $\mathrm{AL}_1(\mathcal{H})$ can be much larger than $\mathrm{L}(\mathcal{H})$ even when $\mathrm{L}(\mathcal{H}) = 1$. For example, for the class of singletons $\mathcal{H} = \{x \mapsto \mathbb{1}\{x = a\} : a \in \mathbb{N}\}$, we have that $\mathrm{L}(\mathcal{H}) = 1$ but $\mathrm{AL}_1(\mathcal{H}) = \infty$.

Theorem 19 (Deterministic realizable upper bound for $\mathrm{L}(\mathcal{H}) = 1$) *For any $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ such that $\mathrm{L}(\mathcal{H}) = 1$, there exists a deterministic learner which, under apple tasting feedback, makes at most $1 + 2\sqrt{T}$ mistakes in the realizable setting.*

Proof We will show that Algorithm 4 makes at most $1 + 2\sqrt{T}$ mistakes in the realizable setting under apple tasting feedback after tuning r.

Let $S = (x_1, h^\star(x_t)), ..., (x_T, h^\star(x_T))$ be the stream observed by the learner, where $h^\star \in \mathcal{H}$ is the optimal hypothesis. As in the proof of Lemma 13, consider splitting the stream into the following three parts. Let S_1 denote those rounds where $\mathrm{L}(V_t^0) = 0$ but $y_t = 0$. Let S_2 denote the rounds where $\mathrm{L}(V_t^0) = 1$, $\hat{y}_t = 1$, but $y_t = 0$. Finally, let S_3 denote the rounds where $\mathrm{L}(V_t^0) = 1$, $\hat{y}_t = 0$,

Algorithm 4 Deterministic Realizable Algorithm For Apple Tasting

Input: $V_1 = \mathcal{H}$ and $r > 0$
Initialize: $C(h) = 0$ for all $h \in \mathcal{H}$
for $t = 1, ..., T$ **do**
  Receive example x_t
  For each $y \in \{0, 1\}$, define $V_t^y = \{h \in V_t \mid h(x_t) = y\}$.
  if $V_t(x_t) = \{y\}$ **then**
    Predict $\hat{y}_t = y$
  else if $\mathrm{L}(V_t^0) = 0$ **then**
    Predict $\hat{y}_t = 1$
    Observe true label y_t
    Update $V_{t+1} = V_t^{y_t}$
  else if $\exists h \in V_t^1$ *such that* $C(h) \geq r$ **then**
    Predict $\hat{y}_t = 1$
    Observe true label y_t
    Update $V_{t+1} = V_t^{y_t}$
  else
    Predict $\hat{y}_t = 0$
    for $h \in V_t^1$ **do**
      Update $C(h)$ += 1
    end
    Set $V_{t+1} = V_t$
end

but $y_t = 1$. The number of mistakes Algorithm 4 makes on the stream S is at most $|S_1|+|S_2|+|S_3|$. We now upper bound each of these terms separately.

Starting with S_1, observe that if $\mathrm{L}(V_t^0) = 0$, then $|V_t^0| \leq 1$. Thus, if $y_t = 0$, Algorithm 4 correctly identifies the hypothesis labeling the data stream and does not make any further mistakes. Accordingly, we have that $|S_1| \leq 1$.

Next, $|S_2|$ is at most the number of times that Algorithm 4 predicts 1 when $\mathrm{L}(V_t^0) = 1$. Note that if $\mathrm{L}(V_t^0) = 1$ then $|V_t^1| \leq 1$. Thus, by the end of the game, there can be at most $\frac{|\{t:\mathrm{L}(V_t^0)=1\}|}{r}$ hypothesis $h \in \mathcal{H}$ such that $C(h) \geq r$. Since Algorithm 4 only predicts 1 when there exists a hypothesis in V_t^1 with count at least r, we have that $|S_2| \leq \frac{|\{t:\mathrm{L}(V_t^0)=1\}|}{r} \leq \frac{T}{r}$.

Finally, we claim that $|S_3| \leq r$. Suppose for the sake of contradiction that $|S_3| \geq r + 1$. Then, by definition, there exists $r+1$ rounds where $\mathrm{L}(V_t^0) = 1$, $\hat{y}_t = 0$ but $y_t = 1$. However, if $\mathrm{L}(V_t^0) = 1$ and $y_t = 1$, then $V_t^1 = \{h^\star\}$. Therefore, on the $r + 1$'th round where $\mathrm{L}(V_t^0) = 1$, $\hat{y}_t = 0$, and $y_t = 1$, it must be the case $C(h^\star) \geq r$. However, if this were true, then the Algorithm would have predicted $\hat{y}_t = 1$ on the $r + 1$'th round, a contradiction. Thus, it must be the case that $|S_3| \leq r$.

Putting it all together, Algorithm 4 makes at most $1 + \frac{T}{r} + r$ mistakes. Picking $r = \sqrt{T}$, gives the mistake bound $1 + 2\sqrt{T}$, completing the proof. ■

We highlight that Theorem 19 is tight up to constants factors. Indeed, for the class $\mathcal{H}$ of singletons over $\mathbb{N}$, we have that $\mathrm{W}(\mathcal{H}) = 2$. Therefore, Theorem 9 implies the lower bound of $\frac{\sqrt{T}}{8}$.

Appendix B. Proof of Lemma 8

To see (i), observe that given any shattered AL tree $\mathcal{T}$ of depth d and width $w_2 > w_1$, we can truncate paths with more than w_1 ones to get a shattered AL tree $\mathcal{T}'$ of the same depth where now every path has at most w_1 ones and the right most path has exactly w_1 ones.

To see (ii), consider the case where $w \leq \mathrm{L}(\mathcal{H})$. Then, by property (i), we have that $\mathrm{AL}_w(\mathcal{H}) \geq \mathrm{AL}_{\mathrm{L}(\mathcal{H})}(\mathcal{H}) \geq \mathrm{L}(\mathcal{H}) \geq w$. If $w > \mathrm{L}(\mathcal{H})$, then $\mathrm{AL}_w(\mathcal{H}) \geq \mathrm{L}(\mathcal{H})$ which follows from the fact that an AL tree $\mathcal{T}$ of width w and depth $\mathrm{L}(\mathcal{H}) < w$ is a complete binary tree of depth $\mathrm{L}(\mathcal{H})$.

To see (iii), fix $w \geq \mathrm{L}(\mathcal{H}) + 1$. Then, by property (ii), we have that $\mathrm{AL}_w(\mathcal{H}) \geq \mathrm{L}(\mathcal{H})$. Thus, it suffices to show that $\mathrm{AL}_w(\mathcal{H}) \leq \mathrm{L}(\mathcal{H})$. Suppose for the sake of contradiction that $\mathrm{AL}_w(\mathcal{H}) \geq \mathrm{L}(\mathcal{H}) + 1$. Then, using property (i) and the fact that $w \geq \mathrm{L}(\mathcal{H}) + 1$, we have that $\mathrm{AL}_{\mathrm{L}(\mathcal{H})+1}(\mathcal{H}) \geq \mathrm{AL}_w(\mathcal{H}) \geq \mathrm{L}(\mathcal{H})+1$. Thus, by definition of ALdim, there exists a Littlestone tree of depth $\mathrm{L}(\mathcal{H})+1$ shattered by $\mathcal{H}$, a contradiction.

To see (iv), note that when $\mathrm{L}(\mathcal{H}) < \infty$, we have that $\mathrm{AL}_{\mathrm{L}(\mathcal{H})+1}(\mathcal{H}) = \mathrm{L}(\mathcal{H})$ by property (iii). Thus, by definition of the Effective width, it must be the case that $\mathrm{W}(\mathcal{H}) \leq \mathrm{L}(\mathcal{H}) + 1$.

To see (v), it suffices to prove that $\mathrm{L}(\mathcal{H}) = \infty \implies \mathrm{W}(\mathcal{H}) = \infty$ since (iv) shows that $\mathrm{L}(\mathcal{H}) < \infty \implies \mathrm{W}(\mathcal{H}) < \infty$. This is true because if $\mathrm{L}(\mathcal{H}) = \infty$, then for any width $w \in \mathbb{N}$ and depth $d \in \mathbb{N}$, one can always prune a shattered Littlestone tree of depth d to get a shattered AL tree of depth d and width w.

Appendix C. Proof of Lemma 13

Algorithm 5 Conversion of Full-Information Algorithm to Apple Tasting Algorithm

Input: Full-Information Algorithm $\mathcal{A}$, false negative mistake bound M_- of $\mathcal{A}$
for $t = 1, ..., T$ **do**
  Receive x_t and query $\mathcal{A}$ to get $\xi_t = \mathcal{A}(x_t)$.
  Draw $r \sim \mathrm{Unif}([0, 1])$ and predict

$$\hat{y}_t = \begin{cases} 1 & \text{if } \xi_t = 1. \\ 1 & \text{if } \xi_t = 0 \text{ and } r \leq \sqrt{M_-/T}. \\ 0 & \text{otherwise.} \end{cases}$$

  If $\hat{y}_t = 1$, receive y_t and update $\mathcal{A}$ by passing (x_t, y_t).
end

If $T \leq M_-$, then the claimed expected mistake bound is $\geq T$, which trivially holds for any algorithm. So, we only consider the case when $T > M_-$. Let $\mathcal{A}$ be a deterministic online learner, which makes at most M_- false negative mistakes and at most M_+ false positive mistakes under full-information feedback. We now show that Algorithm 5, a randomized algorithm that uses $\mathcal{A}$ in a black-box fashion, has expected mistake bound at most $M_+ + 2\sqrt{TM_-}$ in the realizable setting under apple-tasting feedback.

For each bitstring $b \in \{0, 1\}^3$, define $S_b = \{t \in [T] \mid b_1 = \xi_t, b_2 = \hat{y}_t, \text{and } b_3 = y_t\}$. Here, b_1, b_2, b_3 are the first, second, and third bits of the bitstring b. Using this notation, we can write the

expected mistake bound of Algorithm 5 as

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{\hat{y}_t \neq y_t\}\right] = \mathbb{E}\left[\,|S_{101}| + |S_{001}| + |S_{110}| + |S_{010}|\,\right].$$

Since $\hat{y}_t = 1$ whenever $\xi_t = 1$, we have $|S_{101}| = 0$. Note that $|S_{001}| \leq N$, where N is the number of failures before M_- successes in independent Bernoulli trials with probability $\sqrt{M_-/T}$ of success. That is, N quantifies the number of rounds before ξ_t is flipped M_- number of times from 0 to 1 in rounds when $y_t = 1$. Recalling that $N \sim$ Negative-Binomial$(M_-, \sqrt{M_-/T})$, we have

$$\mathbb{E}[|S_{001}|] \leq \mathbb{E}[N] \leq M_-\left(\sqrt{\frac{T}{M_-}} - 1\right) \leq \sqrt{M_-T} - M_-.$$

Moreover, using the fact that $\mathcal{A}$ makes at most M_+ false positive mistakes, we have $|S_{110}| \leq M_+$.

Finally, using the prediction rule in Algorithm 5, we have

$$\mathbb{E}[|S_{010}|] \leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{\xi_t = 0 \;\text{ and }\; \hat{y}_t = 1\}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left\{r \leq \sqrt{\frac{M_-}{T}}\right\}\right] \leq T\sqrt{\frac{M_-}{T}} = \sqrt{M_-T}.$$

Putting everything together, we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{\hat{y}_t \neq y_t\}\right] \leq \sqrt{M_-T} - M_- + M_+ + \sqrt{M_-T} \leq M_+ + 2\sqrt{M_-T}.$$

This completes our proof.

Appendix D. Proof of Lemma 17

Observe that $\hat{\ell}_t(y) \leq \frac{1}{\eta}$ for $y \in \{0,1\}$ since $p_t^1 \geq \eta$. Let $\bar{\ell}_t = \sum_{y\in\{0,1\}} p_t^y \hat{\ell}_t(y)$ and define ℓ'_t such that $\ell'_t(y) = \hat{\ell}_t(y) - \bar{\ell}_t$ for all $y \in \{0,1\}$. Notice that executing EXP4.AT on the loss vectors $\hat{\ell}_1, ..., \hat{\ell}_T$ is equivalent to executing EXP4.AT on the loss vectors $\ell'_1, ..., \ell'_T$. Indeed, since $\bar{\ell}_t$ is constant over the experts, the weights q_t^i remained unchanged regardless of whether ℓ'_t or $\hat{\ell}_t$ is used to update the experts. Moreover, we have that $\ell'_t(y) \geq -\frac{1}{\eta}$.

We start by following the standard analysis of exponential weighting schemes. Let $w_1^i = 1$, $w_{t+1}^i = w_t^i \exp(-\eta \ell_t'(\mathcal{E}_t^i))$, and $W_t = \sum_{i=1}^N w_t^i$. Then, $q_t^i = \frac{w_t^i}{W_t}$ and we have

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{i=1}^N \frac{w_{t+1}^i}{W_t} \\
&= \sum_{i=1}^N \frac{w_t^i \exp(-\eta \ell_t'(\mathcal{E}_t^i))}{W_t} \\
&= \sum_{i=1}^N q_t^i \exp(-\eta \ell_t'(\mathcal{E}_t^i)) \\
&\le \sum_{i=1}^N q_t^i \left(1 - \eta(\ell_t'(\mathcal{E}_t^i)) + \eta^2 (\ell_t'(\mathcal{E}_t^i))^2\right) \\
&= 1 - \eta \sum_{i=1}^N q_t^i \ell_t'(\mathcal{E}_t^i) + \eta^2 \sum_{i=1}^N q_t^i (\ell_t'(\mathcal{E}_t^i))^2,
\end{aligned}$$

where the inequality follows from the fact that $\ell_t'(\mathcal{E}_t^i) \ge -\frac{1}{\eta}$ and $e^x \le 1 + x + x^2$ for all $x \le 1$. Taking logarithms, summing over t, and using the fact that $\ln(1-x) \le -x$ for all $x \ge 0$ we get

$$\ln \frac{W_{T+1}}{W_1} \le -\eta \sum_{t=1}^T \sum_{i=1}^N q_t^i \ell_t'(\mathcal{E}_t^i) + \eta^2 \sum_{t=1}^T \sum_{i=1}^N q_t^i (\ell_t'(\mathcal{E}_t^i))^2.$$

Also, for any expert $j \in [N]$, we have

$$\ln \frac{W_{T+1}}{W_1} \ge \ln \frac{w_{T+1}^j}{W_1} = -\eta \sum_{t=1}^T \ell_t'(\mathcal{E}_t^j)) - \ln N.$$

Combining this with the upper bound on $\ln \frac{W_{T+1}}{W_1}$, rearranging, and dividing by η, we get

$$\sum_{t=1}^T \sum_{i=1}^N q_t^i \ell_t'(\mathcal{E}_t^i) \le \sum_{t=1}^T \ell_t'(\mathcal{E}_t^j) + \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^N q_t^i (\ell_t'(\mathcal{E}_t^i))^2.$$

Using the definition of ℓ_t', we further have that

$$\sum_{t=1}^T \sum_{i=1}^N q_t^i \hat{\ell}_t(\mathcal{E}_t^i) \le \sum_{t=1}^T \hat{\ell}_t(\mathcal{E}_t^j) + \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^N q_t^i (\ell_t'(\mathcal{E}_t^i))^2.$$

Next, observe that

$$\sum_{i=1}^N q_t^i \hat{\ell}_t(\mathcal{E}_t^i) = \left(\sum_{i=1}^N q_t^i \mathcal{E}_t^i\right) \hat{\ell}_t(1) + \left(1 - \sum_{i=1}^N q_t^i \mathcal{E}_t^i\right) \hat{\ell}_t(0) = \frac{1}{1-\eta} \sum_{y \in \{0,1\}} p_t^y \hat{\ell}_t(y) - \frac{\eta}{1-\eta} \hat{\ell}_t(1).$$

Moreover,

$$\begin{aligned}
\sum_{i=1}^{N} q_t^i (\ell_t'(\mathcal{E}_t^i))^2 &= \sum_{i=1}^{N} q_t^i \left(\sum_{y\in\{0,1\}} \mathbb{1}\{y = \mathcal{E}_t^i\} \ell_t'(y) \right)^2 \\
&= \sum_{i=1}^{N} q_t^i \left(\sum_{y\in\{0,1\}} \mathbb{1}\{y = \mathcal{E}_t^i\} \ell_t'(y)^2 \right) \\
&= \sum_{y\in\{0,1\}} \left(\sum_{i=1}^{N} q_t^i \mathbb{1}\{y = \mathcal{E}_t^i\} \right) \ell_t'(y)^2 \\
&= \left(\sum_{i=1}^{N} q_t^i \mathcal{E}_t^i \right) \ell_t'(1)^2 + \left(1 - \sum_{i=1}^{N} q_t^i \mathcal{E}_t^i \right) \ell_t'(0)^2 \\
&\leq \frac{1}{1-\eta} \sum_{y\in\{0,1\}} p_t^y \ell_t'(y)^2.
\end{aligned}$$

Therefore, for any fixed expert j,

$$\frac{1}{1-\eta} \sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \hat{\ell}_t(y) - \frac{\eta}{(1-\eta)} \sum_{t=1}^{T} \hat{\ell}_t(1) \leq \sum_{t=1}^{T} \hat{\ell}_t(\mathcal{E}_t^j) + \frac{\ln N}{\eta} + \frac{\eta}{1-\eta} \sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \ell_t'(y)^2.$$

Multiplying by $1-\eta$ and rearranging, we have

$$\sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \hat{\ell}_t(y) - (1-\eta) \sum_{t=1}^{T} \hat{\ell}_t(\mathcal{E}_t^j) \leq \frac{(1-\eta)\ln N}{\eta} + \eta \sum_{t=1}^{T} \hat{\ell}_t(1) + \eta \sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \ell_t'(y)^2$$

which further implies the guarantee:

$$\begin{aligned}
\sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \hat{\ell}_t(y) - \sum_{t=1}^{T} \hat{\ell}_t(\mathcal{E}_t^j) &\leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \hat{\ell}_t(1) + \eta \sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y \ell_t'(y)^2 \\
&= \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \hat{\ell}_t(1) + \eta \sum_{t=1}^{T} \sum_{y\in\{0,1\}} p_t^y (\hat{\ell}_t(y) - \bar{\ell}_t)^2.
\end{aligned}$$

Next, note that

$$\begin{aligned}\sum_{y\in\{0,1\}} p_t^y(\hat{\ell}_t(y)-\bar{\ell}_t)^2 &= \sum_{y\in\{0,1\}} p_t^y\hat{\ell}_t(y)^2 - \left(\sum_{y\in\{0,1\}} p_t^y\hat{\ell}_t(y)\right)^2 \\ &\le \sum_{y\in\{0,1\}} p_t^y\hat{\ell}_t(y)^2 - \sum_{y\in\{0,1\}} (p_t^y)^2\hat{\ell}_t(y)^2 \\ &= \sum_{y\in\{0,1\}} p_t^y(1-p_t^y)\hat{\ell}_t(y)^2 \\ &\le p_t^1(1-p_t^1)\hat{\ell}_t(0)^2 + p_t^1\hat{\ell}_t(1)^2,\end{aligned}$$

where the first inequality is true because of the nonnegativity of the losses $\hat{\ell}_t$ and the last inequality is true because $0 \le p_t^1 \le 1$. Putting things together, we have that

$$\sum_{t=1}^T \sum_{y\in\{0,1\}} p_t^y\hat{\ell}_t(y) - \sum_{t=1}^T \hat{\ell}_t(\mathcal{E}_t^j) \le \frac{\ln N}{\eta} + \eta\sum_{t=1}^T \hat{\ell}_t(1) + \eta\sum_{t=1}^T p_t^1(1-p_t^1)\hat{\ell}_t(0)^2 + \eta\sum_{t=1}^T p_t^1\hat{\ell}_t(1)^2.$$

Since expert $j \in [N]$ was arbitrary, this completes the proof.

Appendix E. Proof of Theorem 16

From Lemma 17, we have that for an fixed expert $j \in [N]$

$$\sum_{t=1}^T \sum_{y\in\{0,1\}} p_t^y\hat{\ell}_t(y) - \sum_{t=1}^T \hat{\ell}_t(\mathcal{E}_t^j) \le \frac{\ln N}{\eta} + \eta\sum_{t=1}^T \hat{\ell}_t(1) + \eta\sum_{t=1}^T p_t^0 p_t^1\hat{\ell}_t(0)^2 + \eta\sum_{t=1}^T p_t^1\hat{\ell}_t(1)^2.$$

Taking expectations on both sides and using the fact that $\mathbb{E}_t\left[\hat{\ell}_t(y)\right] = \mathbb{1}\{y \ne y_t\}, \mathbb{E}_t\left[\hat{\ell}_t(y)^2\right] = \frac{\mathbb{1}\{y\ne y_t\}}{p_t^1}$ gives

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^T \mathbb{1}\{\hat{y}_t \ne y_t\}\right] - \sum_{t=1}^T \mathbb{1}\{\mathcal{E}_t^j \ne y_t\} &\le \frac{\ln N}{\eta} + \eta\sum_{t=1}^T \mathbb{1}\{1\ne y_t\} + \eta\sum_{t=1}^T \mathbb{E}\left[p_t^0 p_t^1 \frac{\mathbb{1}\{0\ne y_t\}}{p_t^1} + p_t^1\frac{\mathbb{1}\{1\ne y_t\}}{p_t^1}\right] \\ &\le \frac{\ln N}{\eta} + 2\eta T.\end{aligned}$$

Substituting $\eta = \sqrt{\frac{\ln N}{2T}}$, we have

$$\mathbb{E}\left[\sum_{t=1}^T \mathbb{1}\{\hat{y}_t \ne y_t\}\right] - \sum_{t=1}^T \mathbb{1}\{\mathcal{E}_t^j \ne y_t\} \le 2\sqrt{2T\ln N} \le 3\sqrt{T\ln N},$$

which completes the proof.

Appendix F. Proof of Theorem 15

Let $(x_1, y_1), ..., (x_T, y_T)$ denote the stream of labeled instances to be observed by the agnostic learner and let $h^\star = \arg\min_{h \in \mathcal{H}} \sum_{t=1}^{T} \mathbb{1}\{h(x_t) \neq y_t\}$ be the optimal hypothesis in hindsight. Given the time horizon T, let $L_T = \{L \subset [T] : |L| \leq \mathrm{L}(\mathcal{H})\}$ denote the set of all possible subsets of $[T]$ of size of $\mathrm{L}(\mathcal{H})$. For every $L \in L_T$, define an expert $\mathcal{E}_L$, whose prediction on time point $t \in [T]$ on instance x_t is defined by

$$\mathcal{E}_L(x_t) = \begin{cases} \text{SOA}\left(x_t | \{(x_i, \mathcal{E}_L(x_i))\}_{i=1}^{t-1}\right), & \text{if } t \notin L \\ \neg\text{SOA}\left(x_t | \{(x_i, \mathcal{E}_L(x_i))\}_{i=1}^{t-1}\right), & \text{otherwise} \end{cases}$$

where $\text{SOA}\left(x_t | \{(x_i, \mathcal{E}_L(x_i))\}_{i=1}^{t-1}\right)$ denotes the prediction of the SOA on the instance x_t after running and updating on the labeled stream $\{(x_i, \mathcal{E}_L(x_i))\}_{i=1}^{t-1}$. Let $E = \{\mathcal{E}_L : L \in L_T\}$ denote the set of all Experts parameterized by subsets $L \in L_T$. Observe that $|E| \leq T^{\mathrm{L}(\mathcal{H})}$.

We claim that there exists an expert $\mathcal{E}_{L^\star} \in E$ such that for all $t \in [T]$, we have that $\mathcal{E}_{L^\star}(x_t) = h^\star(x_t)$. To see this, consider the hypothetical stream of instances labeled by the optimal hypothesis $S^\star = \{(x_t, h^\star(x_t))\}_{t=1}^{T}$. Let $L^\star = \{t_1, t_2, ...\}$ be the indices on which the SOA would have made mistakes had it run and updated on S^*. By the guarantee of SOA, we know that $|L^\star| \leq \mathrm{L}(\mathcal{H})$. By construction of E, there exists an expert $\mathcal{E}_{L^\star}$ parameterized by $L^\star$. We claim that for all $t \in [T]$, we have that $\mathcal{E}_{L^\star}(x_t) = h^\star(x_t)$. This follows by strong induction on $t \in [T]$. For the base case $t = 1$, there are two subcases to consider. If $1 \in L^\star$, then we have that $\mathcal{E}_{L^\star}(x_1) = \neg\text{SOA}\left(x_1 | \{\}\right) = h^\star(x_1)$, by definition of $L^\star$. If $1 \notin L^\star$, then $\mathcal{E}_{L^\star}(x_1) = \text{SOA}\left(x_1 | \{\}\right) = h^\star(x_1)$ also by definition of $L^\star$. Now for the induction step, suppose that $\mathcal{E}_{L^\star}(x_i) = h^\star(x_i)$ for all $i \leq t$. Then, if $t+1 \in L^\star$, we have that $\mathcal{E}_{L^\star}(x_{t+1}) = \neg\text{SOA}\left(x_{t+1} | \{(x_i, \mathcal{E}_{L^\star}(x_i))\}_{i=1}^{t}\right) = \neg\text{SOA}\left(x_{t+1} | \{(x_i, h^\star(x_i))\}_{i=1}^{t}\right) = h^\star(x_{t+1})$. If $t+1 \notin L^\star$, then $\mathcal{E}_{L^\star}(x_{t+1}) = \text{SOA}\left(x_{t+1} | \{(x_i, \mathcal{E}_{L^\star}(x_i))\}_{i=1}^{t}\right) = \text{SOA}\left(x_{t+1} | \{(x_i, h^\star(x_i))\}_{i=1}^{t}\right) = h^\star(x_{t+1})$. The final equality in both cases are due to the definition of $L^\star$.

Now, consider the agnostic online learner $\mathcal{A}$ that runs EXP4.AT using E. By Theorem 16, we have that

$$\begin{aligned} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{\mathcal{A}(x_t) \neq y_t\}\right] &\leq \inf_{\mathcal{E} \in E} \sum_{t=1}^{T} \mathbb{1}\{\mathcal{E}(x_t) \neq y_t\} + 3\sqrt{T \ln |E|} \\ &\leq \sum_{t=1}^{T} \mathbb{1}\{\mathcal{E}_{L^\star}(x_t) \neq y_t\} + 3\sqrt{\mathrm{L}(\mathcal{H}) T \ln T} \\ &= \sum_{t=1}^{T} \mathbb{1}\{h^\star(x_t) \neq y_t\} + 3\sqrt{\mathrm{L}(\mathcal{H}) T \ln T}. \end{aligned}$$

Thus, $\mathcal{A}$ achieves the stated upper bound on expected regret under apple tasting feedback, which completes the proof.

Appendix G. Effective width of the k-wise generalization of $\mathcal{H}_{\text{sing}}$

In this section, we compute the Effective width of the k-wise generalization of the class of singletons $\mathcal{H}_{\text{sing}}$.

Proposition 20 *Let* $\mathcal{X} = \mathbb{N}$ *and* $\mathcal{H}_k = \{x \mapsto \mathbb{1}\{x \in A\} : A \subset \mathbb{N}, |A| \leq k\}$. *Then,* $\mathrm{W}(\mathcal{H}) = k+1$ *and* $\mathrm{AL}_{\mathrm{W}(\mathcal{H})} = 0$.

Proof Consider an Apple Littlestone tree $\mathcal{T}(w, d)$ of width $w = k$ and depth $d \geq w$ such that all the internal nodes on level $i \in [d]$ are labeled by the instance $i \in \mathbb{N}$. It is not too hard to see that $\mathcal{H}_k$ shatters $\mathcal{T}(w, d)$. Since $d \geq w$ was chosen arbitrarily, this is true for all $d \in ,N$ and thus $\mathrm{AL}_k(\mathcal{H}_k) = \infty$. On the other hand, consider an Apple Littlestone tree $\mathcal{T}'(w', d)$ of width $w' = k+1$ and depth $d \in \mathbb{N}$. Note that in order to shatter $\mathcal{T}'$, there must exist a hypothesis that outputs at least $k + 1$ ones across $k + 1$ distinct instances in $\mathcal{X}$. However, by definition, every hypothesis $h \in \mathcal{H}$ outputs 1 on at most k distinct instances. Thus, $\mathcal{T}'$ cannot be shattered by $\mathcal{H}_{k+1}$. Since this is true for all such $d \in \mathbb{N}$, we have that $\mathrm{AL}_{k+1}(\mathcal{H}_k) = 0$. This completes the proof as it must be the case that $\mathrm{W}(\mathcal{H}) = k + 1$. ■