Universal Rates for Multiclass Learning with Bandit Feedback

Steve Hanneke STEVE.HANNEKE@GMAIL.COM
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Amirreza Shaeiri AMIRREZA.SHAEIRI@GMAIL.COM
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Qian Zhang ZHAN3761@PURDUE.EDU
Department of Statistics, Purdue University, West Lafayette, IN, USA

Editors: Nika Haghtalab and Ankur Moitra

Abstract

The seminal work of Daniely, Sabato, Ben-David, and Shalev-Shwartz (2011) introduced the problem of multiclass learning under bandit feedback and provided a combinatorial characterization of its learnability within the framework of PAC learning. In multiclass learning under bandit feedback, there is an unknown data distribution over an instance space $\mathcal{X}$ and a (possibly infinite) label space $\mathcal{Y}$ similar to classical multiclass learning, but the learner does not directly observe the correct labels of the i.i.d. training examples. Instead, during each round, the learner receives an example, makes a prediction for its label, and receives bandit feedback only indicating whether the prediction is correct. Despite this restriction, the goal remains the same as in classical multiclass learning, where the objective is to output a function that correctly classifies most future examples generated by the same underlying data distribution.

In the present work, we study the problem of multiclass learning under bandit feedback within the framework of *universal learning* (Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff, 2021). Unlike the classical uniform learning framework, the universal learning framework seeks to establish guarantees that hold for every fixed data distribution, but without requiring these guarantees to be uniform across all distributions. This makes it possible to study the behavior of learning curves. In the *uniform learning* framework, no concept class $\mathcal{C}$ is learnable when the effective label space is unbounded. This arises from the need to guess an unknown natural number while keeping the number of guesses uniformly bounded over the choices of the target number. In contrast, surprisingly, we demonstrate that the universal learnability of concept classes $\mathcal{C}$ even when the effective label space is unbounded gives rise to a rich theory.

More concretely, our primary contribution is a theory that reveals an inherent trichotomy governing instance optimal learning curves in the realizable setting. Moreover, the best achievable universal learning rate for any given concept class can only decay either at an *exponential*, a *linear*, or an *arbitrarily slow* rate. In particular, the trichotomy is combinatorially characterized by the absence of an infinite multiclass Littlestone tree and the combination of an infinite Natarajan Littlestone tree and an infinite progressive Littlestone tree. Furthermore, we introduce novel learning algorithms for achieving instance optimal universal rates.

Keywords: Universal Learning, Bandit Feedback, Multiclass Learning

© 2025 S. Hanneke, A. Shaeiri & Q. Zhang.

1. Introduction

Consider a scenario in which a pandemic is caused by an unknown virus, where we want to develop an AI model to recommend effective medicines to patients. Each patient responds positively to only one specific drug from a predetermined list of medicines. Due to various challenges associated with human clinical trials, including ethical and financial, the number of patients available for testing is limited. In addition, the only feedback that can be obtained is whether a specific tested drug was effective for a given patient. In this context, the primary objective is to find a model that accurately recommends the appropriate medicine for future patients drawn from the same underlying population. This leads to two critical questions: (1) How many patients must be involved in the initial testing phase to ensure high accuracy? (2) What is the algorithm to achieve the guarantee of the previous question?

The above example and similar real-world situations can be formulated in a framework called *Multiclass Learning with Bandit Feedback*. Roughly speaking, in multiclass learning under bandit feedback framework, initially, nature selects an unknown data distribution over an instance space $\mathcal{X}$ (e.g. space of images) and a label space $\mathcal{Y}$ (e.g. categories of images). Fix a sample size $n \in \mathbb{N}$. Subsequently, the learner and nature interact sequentially in n rounds. In particular, during each round $t \in \{1, 2, \ldots, n\}$, nature first independently samples an example (x_t, y_t) from her chosen distribution and reveals the instance x_t to the learner. The learner is then tasked with predicting a label from the label space for the received instance. Upon the learner's prediction, nature only reveals whether the prediction is correct. Eventually, the learner must output a function from instance space to label space that correctly classifies most future examples generated by the same unknown data distribution. Following standard learning theory conventions, we consider a concept class $\mathcal{C}$, consisting of functions mapping the instance space to the label space. Moreover, we focus on the realizable setting, where we assume that samples from the data distribution are almost surely consistent with at least one concept in the concept class.

The mentioned framework may bring to our mind the fundamental problem of contextual bandit in interactive decision making. Michael Woodroofe wrote the first paper on this subject in 1979 (Woodroofe, 1979), where he also used clinical trials as a motivating example. Since then, a growing body of research has been extensively studying this topic under dissimilar names, including bandit problems with side information, associative bandit problems, and bandit problems with a concomitant variable, among others; see the survey article of Tewari and Murphy (2017) and references therein. Notably, beyond clinical trials, this problem has found other practical applications, such as news article recommendations (Li, Chu, Langford, and Schapire, 2010; Agarwal, Bird, Cozowicz, Hoang, Langford, Lee, Li, Melamed, Oshri, Ribas, et al., 2016). Moreover, our framework can be framed as a special instance of the contextual bandit framework. However, reducing it to a general contextual bandit framework overlooks a unique structural property of the reward function in the classification setting, namely its sparsity. Indeed, previous studies have leveraged this sparsity to achieve improved sample complexity bounds (Erez, Cohen, Koren, Mansour, and Moran, 2024b,a).

Going forward, prior research on this problem has only considered the *uniform* learning framework. The seminal work of Daniely, Sabato, Ben-David, and Shalev-Shwartz (2011) introduced the problem of multiclass learning under bandit feedback and provided a combinatorial characterization of its learnability within the framework of PAC learning. Moreover, very recent work of Erez, Cohen, Koren, Mansour, and Moran (2024b,a) continued that line of research by focusing mainly

on the finite size concept classes, trying to improve the dependence of the bounds on the number of labels. In essence, the *uniform* learning framework seeks a theoretical guarantee that is true for all distributions without any dependence on the distribution itself. This raises the following limitations:

- In practice, the performance of a learning algorithm is often evaluated by examining its "learning curve". This plot consists of points that represent the error of the functions outputted by the algorithm, with each being trained on a distinct number of training examples and evaluated on a sufficient amount of unseen examples. Empirical studies have shown that the error decay rate observed in such learning curves for a given learning scenario can significantly exceed the rates predicted by theories based on the uniform learning framework (Cohn and Tesauro, 1990, 1992; Schuurmans, 1997; Viering and Loog, 2022). This is due to the nature of the uniform learning framework: it can only capture an upper envelope of the learning curves over all learning distributions, in contrast with practical results where the distribution is fixed.

- In addition, if the number of labels is unbounded, particularly if there exists at least one instance in the instance space for which concepts in the concept class can take on an unbounded number of distinct labels, uniform learnability is not possible. To see this, fix a learning algorithm $\mathbf{A}$ and a sample size $n \in \mathbb{N}$ such that we can achieve an error rate of at most $1/2$ with a probability of at least $1/2$ for every realizable data distribution. Now, take $x \in \mathcal{X}$ for which concepts in the concept class can take at least $2n + 1$ distinct labels. In particular, let $\mathcal{B}$ be the set of those $2n + 1$ labels. Next, define P as the uniform distribution on $\mathcal{B}$. Indeed, the error rate of $\mathbf{A}$ on the worst-case realizable distribution is at least the expected error rate of it when we choose the data distribution based on P. It is not hard to see that the probability of guessing the correct label after n trials is at most $n/(2n + 1) < 1/2$. As a result, for any $\mathbf{A}$, there exists a realizable distribution such that with probability greater than $1/2$ the error rate of $\mathbf{A}$ is greater than $1/2$. For more details, see Section C. Consequently, in multiclass learning with bandit feedback under the uniform learning framework, no concept class is learnable when the effective label space is unbounded. This includes elementary and natural classes, such as a class of countably infinite collection of constant functions over some domain, which is uniformly learnable with one example in the classical multiclass learning framework.

In response to the above theoretical limitations, we focus on the arguably more realistic framework of *universal* learning introduced in Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). This framework makes it possible to study theoretical guarantees that are still true for all distributions, but it also allows a dependence on the distribution itself. As a result, it can possibly capture optimal asymptotic convergence rates for learning curves. Moreover, let us explore an instance of multiclass learning under the bandit feedback framework involving an instance space with only one element, a label space as natural numbers, and a concept class consisting of all possible functions from the given instance space to the specified label space. Now, consider the following learning algorithm. In each round $i \in \mathbb{N}$, the algorithm predicts the label i; eventually, after traversing all training examples, if it has received any feedback confirming that the prediction is correct, the algorithm outputs the correct function, otherwise it outputs arbitrarily. It is not hard to see that this algorithm can achieve exponential decay in the expected error rate as a function of the number of examples. In the current manuscript, surprisingly, in the multiclass learning under bandit

feedback framework, we demonstrate that universal learnability of concept classes $\mathcal{C}$ even when the effective label space is unbounded gives rise to a rich theory.

1.1. Overview of the Main Results

In the following subsection, we provide a detailed summary of the key results and findings presented in our paper.

1.1.1. MULTICLASS LEARNING UNDER BANDIT FEEDBACK FRAMEWORK

To define multiclass universal learning under bandit feedback framework, we have the following components. Fix a non-empty set $\mathcal{X}$ admitting the minimal set theoretic assumption as the instance space. For example, the instance space $\mathcal{X}$ can be a Euclidean space or any countable set. Also, fix a non-empty and countable set $\mathcal{Y}$ as the label space. Moreover, for the power set σ-algebra $\sigma(\mathcal{Y})$ over the label space $\mathcal{Y}$, define $\Pi(\mathcal{Y})$ to be the set of all probability measures on $(\mathcal{Y}, \sigma(\mathcal{Y}))$. A pair $z = (x, y) \in \mathcal{X} \times \mathcal{Y}$ is called an example, and $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ denotes the space of examples. A concept is a function from $\mathcal{X}$ to $\mathcal{Y}$. With this in mind, fix a non-empty set of concepts $\mathcal{C}$ that satisfies a minimal measurability assumption as the concept class. In particular, a 3-tuple $\mathcal{Q} = \big(\mathcal{X}, \mathcal{Y}, \mathcal{C}\big)$ presents an instance of the multiclass universal learning under bandit feedback framework. Notably, we define a non-degenerate instance of multiclass learning under bandit feedback framework to remove trivial cases. See Section B.2.1 for more details.

A data distribution $\mathcal{D}$ is a probability measure on $\mathcal{Z} = (\mathcal{X} \times \mathcal{Y})$. Initially, nature selects a data distribution $\mathcal{D}$. Fix $n \in \mathbb{N}$ as the number of samples. Then, we have an n-rounded sequential game between the learner and nature. Moreover, in each round $t \in \{1, 2, \ldots, n\}$:

1. Nature samples an instance (x_t, y_t) from $\mathcal{D}$ and reveals x_t to the learner.
2. The learner predicts a label $\hat{y}_t$ from $\mathcal{Y}$.
3. Nature reveals the feedback $f_t = \mathbb{1}\big\{y_t \neq \hat{y}_t\big\}$ to the learner.

Finally, perhaps based on $\{(x_1, \hat{y}_1, f_1), (x_2, \hat{y}_2, f_2), \ldots, (x_n, \hat{y}_n, f_n)\}$, the learner outputs a universally measurable function from $\mathcal{X}$ to $\mathcal{Y}$. In contrast, we call the framework multiclass learning with *full supervision* (or with full-information feedback) if in step 3 the nature reveals the true label y_t to the learner.

Let Σ be defined as $\Sigma := \mathcal{Y} \times \{0, 1\}$. A learning algorithm $\mathbf{A}$ is a 2-tuple $(\mathbf{A}_1, \mathbf{A}_2)$ of mappings: $\mathbf{A}_1$ is a mapping from $(\mathcal{X} \times \Sigma)^\star \times \mathcal{X}$ to a probability measure $\nu \in \Pi(\mathcal{Y})$, and $\mathbf{A}_2$ is also a mapping from $(\mathcal{X} \times \Sigma)^\star \times \mathcal{X}$ to a probability measure $\mu \in \Pi(\mathcal{Y})$. Notably, both $\mathbf{A}_1$ and $\mathbf{A}_2$ are defined for generality to incorporate randomness. We define two separate mappings, one related to the game and one related to the final result, because we believe that this approach is more convenient to understand. In particular, for the learning algorithms constructed in this paper, $\mathbf{A}_1$ may be randomized, but $\mathbf{A}_2$ is deterministic; that is, given a fixed output from $\mathbf{A}_1$, the output of $\mathbf{A}_2$ is deterministic. Meanwhile, all of our lower bounds remain valid up to constant factors for our definition of general learning algorithms, specifically, even if we allow internal randomness in $\mathbf{A}_2$ as well.

Now, let $\mathbf{A}$ be a learning algorithm. For every finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, we denote by $\mathbf{B}_{\mathbf{A}}(\mathcal{S})$ the random variable taking values in $(\mathcal{X} \times \Sigma)^n$ representing the first part of the input of $\mathbf{A}_1$ in round $n+1$ when the choices of nature are obtained according to $\mathcal{S}$. Building upon that, also for every finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, we denote by $\hat{h}^{\mathcal{S}}_{\mathbf{A}}$ the random function taking values in $\mathcal{Y}^{\mathcal{X}}$ such that for every $x \in \mathcal{X}$, we have: $\hat{h}^{\mathcal{S}}_{\mathbf{A}}(x) = \hat{y}$, where $\hat{y} \sim \mathbf{A}_2((\mathbf{B}_{\mathbf{A}}(\mathcal{S}), x))$. Note that whenever it is clear from the context, we may simply write $\hat{h}$ instead of $\hat{h}^{\mathcal{S}}_{\mathbf{A}}$. See Section B.2.3 for more details.

For a data distribution $\mathcal{D}$ over $\mathcal{Z}$ and a concept $c : \mathcal{X} \to \mathcal{Y}$, define the error rate of a concept c with respect to the data distribution $\mathcal{D}$ denoted by $\mathrm{er}_{\mathcal{D}}(c)$ as follows: $\mathrm{er}_{\mathcal{D}}(c) := \mathcal{D}(\{(x, y) \in \mathcal{Z} : c(x) \neq y\})$. In particular, for a data distribution $\mathcal{D}$, a learning algorithm $\mathbf{A}$, and a finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, the error rate of $\hat{h}^{\mathcal{S}}_{\mathbf{A}}$ with respect to the distribution $\mathcal{D}$ is $\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) = \mathcal{D}(\{(x, y) \in \mathcal{Z} : \hat{h}^{\mathcal{S}}_{\mathbf{A}}(x) \neq y\})$. See Section B.2.4 for more details.

Based on the previous definition, we say that a data distribution $\mathcal{D}$ is realizable by a concept class $\mathcal{C}$ if $\inf_{c \in \mathcal{C}} \ \mathrm{er}_{\mathcal{D}}(c) = 0$. Furthermore, we denote by $\mathrm{RE}(\mathcal{C})$ the set of all realizable data distributions. See Section B.2.5 for more details.

A function $\mathcal{R} : \mathbb{N} \to (0, 1]$ is a rate function if $\lim_{n \to \infty} \mathcal{R}(n) = 0$. Moreover, we use the subscript of n to emphasize that we have a rate function. In the current manuscript, we refer to $\mathcal{R}(n) = C\,\mathrm{e}^{-cn}$ for some constants $C, c \in \mathbb{R}^+$ as the **exponential rate**. In addition, we refer to $\mathcal{R}(n) = C/n$ for some constant $c \in \mathbb{R}^+$ as the **linear rate**.

Now, based on the simple definition above, we formalize the objective in multiclass universal learning under bandit feedback framework. See Section B.2.6 for more details.

Definition 1 *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass learning under bandit feedback framework. Also, let $\mathcal{R}$ be a rate function. Then, we say that:*

- *$\mathcal{Q}$ is universally learnable under bandit feedback at rate $\mathcal{R}$, if there exists a learning algorithm $\mathbf{A}$ such that for every realizable distribution $\mathcal{D}$, there exist $C, c \in \mathbb{R}^+$ such that for every $n \in \mathbb{N}$, we have: $\mathbb{E}_{\mathcal{S} \sim \mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \leqslant C\mathcal{R}(cn)$.*
- *$\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}$, if for all learning algorithms $\mathbf{A}$, there exists a realizable distribution $\mathcal{D}$ and $C, c \in \mathbb{R}^+$ such that for infinitely many $n \in \mathbb{N}$, we have: $\mathbb{E}_{\mathcal{S} \sim \mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \geqslant C\mathcal{R}(cn)$.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback at optimal rate $\mathcal{R}$, if $\mathcal{Q}$ is universally learnable under bandit feedback at rate $\mathcal{R}$ and $\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}$.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback, if there exists a learning algorithm $\mathbf{A}$ such that for every realizable distribution $\mathcal{D}$, we have: $\lim_{n \to \infty} \mathbb{E}_{\mathcal{S} \sim \mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] = 0$.*
- *$\mathcal{Q}$ requires arbitrarily slow rates under bandit feedback, if for every rate function $\mathcal{R}'$, we know that $\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}'$.*

1.1.2. OPTIMAL UNIVERSAL LEARNING RATES

Now, we are ready to state the main results. The aim of this paper is to characterize the optimal learning rates achievable by a learning algorithm. Our first main result reveals a fundamental trichotomy of optimal universal learning rates. Formally, we have the following theorem.

Theorem 2 (Trichotomy of Optimal Universal Learning Rates) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of the multiclass learning under bandit feedback framework. Then, exactly one of the following statements holds:*

- *$\mathcal{Q}$ is universally learnable under bandit feedback at the optimal rate e^{-n}.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback at the optimal rate $1/n$.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback, but requires arbitrarily slow rates under bandit feedback.*

We emphasize that all previous work on this setting has considered the PAC learning model. Our next theorem combinatorially characterizes each of the cases in the above theorem. In particular, the combinatorial characterization is based on the absence of an infinite multiclass Littlestone tree and the combination of an infinite Natarajan Littlestone tree and an infinite progressive Littlestone tree.

Having an infinite multiclass Littlestone tree is closely related to having an infinite multiclass Littlestone dimension introduced in Daniely, Sabato, Ben-David, and Shalev-Shwartz (2011). However, it is important to distinguish between the two. An infinite multiclass Littlestone dimension can arise from the presence of finite Littlestone trees with arbitrarily large depth, which does not necessarily imply the existence of a single tree with infinite depth. For more details, refer to Section B.3 and Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). Having an infinite Natarajan Littlestone tree is also closely related to having an infinite Natarajan dimension introduced in the work of Natarajan and Tadepalli (1988). For more details, refer to Section B.3 and Hanneke, Moran, and Zhang (2023); Kalavasis, Velegkas, and Karbasi (2022). Furthermore, our work introduces a novel combinatorial object called the Progressive Littlestone tree. Informally, it is a Littlestone tree that has an increasing number of children, where nodes at depth $d \in \mathbb{N}$ should have $d + 1$ children. Interestingly, having an infinite such a tree implies an infinite List Littlestone tree for any finite list size. List Littlestone trees introduced in Moran, Sharon, Tsubari, and Yosebashvili (2023) on List online learning. Formally, we have the following theorem.

Theorem 3 (Combinatorial Characterization of Optimal Universal Learning Rates) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of the multiclass learning under bandit feedback framework. Then, the followings hold:*

- *If $\mathcal{C}$ does not have an infinite multiclass Littlestone tree, then $\mathcal{Q}$ is universally learnable under bandit feedback at the optimal rate e^{-n}.*

- *If $\mathcal{C}$ has an infinite multiclass Littlestone tree, $\mathcal{C}$ does not have an infinite Natarajan Littlestone tree, and $\mathcal{C}$ does not have an infinite progressive Littlestone tree, then $\mathcal{Q}$ is universally learnable under bandit feedback at the optimal rate $1/n$.*
- *If $\mathcal{C}$ has either an infinite Natarajan Littlestone tree or an infinite progressive Littlestone tree, then $\mathcal{Q}$ is universally learnable under bandit feedback, but requires arbitrarily slow rates under bandit feedback.*

The definitions of $\mathcal{C}$ possessing an infinite multiclass Littlestone tree, an infinite Natarajan Littlestone tree, and an infinite progressive Littlestone tree are provided in Definition 18, Definition 24, and Definition 21, respectively.

The above theorem provides a complete theory for multiclass universal learning rates under bandit feedback. First, it allows us to overcome the inherent limitations of the PAC learning framework previously discussed. Moreover, it enables the learning of classes such as a class of countably infinite collection of constant functions over some domain.

Remark 4 *We have established optimal results in Theorem 2 and Theorem 3. Interestingly, the corresponding optimal results for multiclass learning under full supervision remain unknown, as noted in Hanneke, Moran, and Zhang (2023).*

Remark 5 *In Theorem 3, we showed: if $\mathcal{C}$ has neither an infinite Natarajan Littlestone tree nor an infinite Progressive Littlestone tree, then $\mathcal{Q}$ is universally learnable under bandit feedback at the rate $1/n$. Interestingly, having no infinite Natarajan Littlestone tree is not sufficient for universal learnability of $\mathcal{Q}$ under full supervision at the optimal rate $1/n$. In fact, we have: if $\mathcal{C}$ does not have an infinite DSL tree (Hanneke, Moran, and Zhang, 2023, Definition 7), then $\mathcal{Q}$ is universally learnable under full supervision at the rate $\log(n)/n$. See Hanneke, Moran, and Zhang (2023, 2024b) for more details.*

Remark 6 *Notably, our findings suggest that in multiclass learning under bandit feedback within the framework of universal learning in the realizable setting, knowing the sequence of examples in advance and being able to adaptively determine the order for making predictions is not beneficial. In the field of online learning in Littlestone's setup Littlestone (1988), one can consider problems in which the set of instances is known before the start of the game. Among these, the most flexible is self-directed online learning, where the online learning algorithm is allowed to select the next instance for prediction from the remaining set of instances in each round Goldman et al. (1993); Goldman and Sloan (1994); Ben-David et al. (1995); Ben-David and Eiron (1998); Devulapalli and Hanneke (2024). Specifically, in this new setting of multiclass learning under bandit feedback, after the training sequence is sampled from the underlying distribution, the learner is allowed to adaptively select the next instance from that sequence to predict based on previous decisions and feedback. Indeed, our upper bounds still hold in the above case. On the other hand, the lower bound under the existence of an infinite progressive Littlestone tree remains valid as well. In particular, that lower bound is based on the probability that the number of training instances which coincide with the test instance is not large enough under the constructed hard distribution, and this probability does not depend on the decisions made by the learner. Other lower bounds carry over from the multiclass setting. Therefore, even in the most flexible formulation, the problem remains the same landscape and combinatorial characterization.*

1.2. Overview of the Techniques

In the following subsection, we provide a summary of the key techniques in the proof of the above theorems, namely Theorem 2 and Theorem 3.

1.2.1. LOWER BOUNDS

First, we prove that the finiteness of the effective label space is necessary for multiclass PAC learnability with bandit feedback. The idea behind this proof was explained in the second bullet point in the introduction. Next, we draw on three theorems from Hanneke, Moran, and Zhang (2023) on multiclass universal learnability under full supervision: the lower bound for the exponential rate, the lower bound for the linear rate with an infinite multiclass Littlestone tree, and the lower bound for arbitrarily slow rates with an infinite DSL tree. Indeed, any lower bound established in multiclass universal learning under full supervision should also be applicable to the bandit setting, and an infinite Natarajan Littlestone tree is an infinite DSL tree.

The novel contribution of the manuscript is the lower bound based on the existence of an infinite progressive Littlestone tree. Given an infinite progressive Littlestone tree, we first sample an infinite path $\mathscr{P}$ emanating from the root by choosing a child uniformly at random from all children of the current node and proceed to the chosen child to repeat the above step inductively. A random example is generated by sampling a random number K from some discrete distribution D over $\mathbb{N}$ and picking the labels of the node and its edge at layer K in the infinite path $\mathscr{P}$. For $n+1$ independent random examples, we let the first n examples be training examples and the last example be the test example. Then, if the test example is in layer $k \in \mathbb{N}$ and all training example are at layers at most k among which at most $\lceil k/2 \rceil$ are exactly k (denote this event by $\mathcal{E}_k$), then since the path $\mathscr{P}$ chooses an edge uniformly at random from $k+1$ edges at layer k, with bandit feedback, the error rate of any algorithm is at least $(k - \lceil k/2 \rceil - 1)/(k+1) \geqslant 1/6$ for $k \geqslant 4$. Then, we need to lower bound the probability of $\mathcal{E}_k$. Conditional on all training examples being at layers at most k, the number of those at layer k follows a binomial distribution whose parameter is determined by D. Then, we can tune D so that applying a Chernoff bound, the conditional probability that at most $\lceil k/2 \rceil$ training examples are at layer k is at least a constant. Note that after such tuning, the probability masses of D, layer sequence, and sample size sequence differ by some multiplicative constant from those in Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). See Section C for more details.

1.2.2. EXPONENTIAL UPPER BOUND

First, we very briefly explain the approach of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). Their algorithm that achieves the exponential rate if the binary concept class has no infinite Littlestone tree utilizes a result on universal online learning as the first piece. Next, given a sample $\mathcal{S}$ of size n, they estimate a batch size parameter $t^\star$. Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint subsets of $\mathcal{S}$ each of size $\lfloor n/2 \rfloor$. In particular, based on their approach, for each batch size t between 1 and $\lfloor n/2 \rfloor$, we have $\lfloor n/2t \rfloor$ predictors each trained on a disjoint batch of size t from $\mathcal{A}$. Then, for each batch size t between 1 and $\lfloor n/2 \rfloor$, we run a test based on the errors of the predictors resulting from the previous sentence on all members of $\mathcal{B}$. Finally, we choose the smallest batch size t that passes

the test as $\hat{t}$. Based on the mentioned estimation, they use the majority vote of the predictors each trained on the disjoint batch of size $\hat{t}$ from $\mathcal{A}$ to derive the final predictor.

Now, we draw the first piece from the recent work of Hanneke, Shaeiri, and Wang (2024c). In particular, they established a similar result as in Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021) on universal multiclass online learning under bandit feedback. Given the correct estimation of the batch size, the last step is also similar to that of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). However, generally speaking, in the bandit setting, we may not reuse samples because of the sequential nature of the problem. To address this, we introduce two novel ideas. First, for every batch size t from 1 to $\lfloor\sqrt{n}\rfloor$, we take two subsets of the samples $\mathcal{A}_t$ and $\mathcal{B}_t$ each of size $\lfloor n/\big(2i(i+1)\big)\rfloor$, so that all subsets are disjoint. The subset $\mathcal{A}_t$ is used to construct batches of size t, while $\mathcal{B}_t$ is used to evaluate the resulting predictors, similar to what we explain in the previous paragraph based on Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). However, unlike Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021), we may still be unable to reuse $\mathcal{B}_i$ for the tests. To handle this, we modify the test criterion. Specifically, for every batch size t from 1 to $\lfloor\sqrt{n}\rfloor$, we require that the majority vote of the resulting functions based on $\mathcal{A}_t$ to be completely correct on $\mathcal{B}_t$. Additionally, we incorporate a similar criterion to that of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). In fact, we require that both criterion should hold. If the first criterion is satisfied, it allows us to fully label $\mathcal{B}_t$ and subsequently verify the second criterion. We then demonstrate how to adapt this modified criterion within the proof of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). See Section D for more details.

1.2.3. LINEAR UPPER BOUND

Kalavasis, Velegkas, and Karbasi (2022) constructed an algorithm achieving a linear rate if $\mathcal{C}$ has no infinite Natarajan Littlestone (NL) tree for finite label space under full supervision. To incorporate their approach in our setting, we consider to reduce the label space by learning a finite list and acquire sufficient correctly labeled training examples by random guessing. Let us first assume full supervision. We can relate the progressive Littlestone tree to an adversarial game between two players $\mathtt{P}_A$ and $\mathtt{P}_L$, where in round $t \in \mathbb{N}$, informally, $\mathtt{P}_A$ chooses an instance $x_t \in \mathcal{X}$ and a list of $t+1$ distinct labels in $\mathcal{Y}$, and then $\mathtt{P}_L$ picks a label y_t from the list. $\mathtt{P}_L$ wins the game in round t if $((x_i, y_i))_{i=1}^t$ is not $\mathcal{C}$-realizable. If $\mathcal{C}$ has no infinite progressive Littlestone tree, $\mathtt{P}_L$ has a universally measurable winning strategy. Using this strategy, we can construct an online list learning algorithm which outputs a menu given training examples, where a menu μ is a function that maps an instance in $\mathcal{X}$ to a subset of $\mathcal{Y}$, its size is $|\mu| := \sup_{x\in\mathcal{X}} |\mu(x)|$, and its error rate under distribution P on $\mathcal{Z}$ is $P(\{(x,y) \in \mathcal{Z} : y \notin \mu(x)\})$. We show that the probability that this algorithm outputs a menu with zero error rate and bounded size converges to 1 as the size of training samples goes to infinity. To apply this algorithm with bandit feedback, we have to obtain the true labels for some training examples. This is achieved by guessing through sampling a label independently for each instance from a common distribution D over $\mathcal{Y}$ so that the probability of guessing correctly is positive. Since $\mathcal{Y}$ is countable, we can simply pick a distribution D with full support on $\mathcal{Y}$. Then, we can show that for an infinite sequence of i.i.d. samples from data distribution P over $\mathcal{Z}$, those whose labels are guessed correctly form an infinite sequence of i.i.d. samples from a distribution P_D defined by $P_D(E) := \mathbb{P}_{((X,Y),Y')\sim P\times D}((X,Y) \in E \mid Y = Y')$ for any measurable set $E \subseteq \mathcal{Z}$. Though P_D is different from P, by the full supportedness of D, we prove that $P_D \in \mathrm{RE}(\mathcal{C})$ if $P \in \mathrm{RE}(\mathcal{C})$ and

$P(E) > 0$ implies $P_D(E) > 0$ for all measurable $E \subseteq \mathcal{Z}$. With those properties, we can extend the theoretical guarantee of the list learning algorithm from the full supervision setting to the bandit feedback setting. Applying this algorithm to $m = \lfloor \sqrt{n/2} \rfloor$ batches each containing m examples to create m menus and aggregating them properly, we obtain a menu $\hat{\mu}$ of bounded size and zero error rate with high probability.

Using those examples from the first half of the training sequence whose labels are guessed correctly as in the above procedure, since $\mathcal{C}$ has no infinite NL tree, we can run the algorithm from Kalavasis, Velegkas, and Karbasi (2022) based on playing an adversarial game related to the NL tree to construct several concept classes with bounded Natarajan dimension. We show that with high probability, a majority proportion of those classes contain the sequence of the true labels from the second half of the training examples and the test example. Then, we modify those concept classes by excluding those concepts which are not consistent with $\hat{\mu}$ from the class. Since the size of $\hat{\mu}$ is bounded, those modified concept classes can be treated as having finite label space. Now, we can apply the one-inclusion algorithm (Algorithm 1) with those concept classes on the second half of the training examples to produce a classifier for each concept class. The final classifier is the majority vote of those output classifiers. Applying the linear rate of the one-inclusion algorithm using concept classes having finite Natarajan dimension and finite label space, we can upper bound the expected error rate of the majority vote based on Markov's inequality. See Section E for details.

1.2.4. UNIVERSAL CONSISTENCY

Hanneke, Kontorovich, Sabato, and Weiss (2021) constructed a universally consistent (i.e., its expected error rate converges to zero as the training sample size goes to infinity for all realizable distributions) multiclass learning algorithm $\mathcal{A}$ under full supervision. To apply it under the bandit feedback framework, just as we do in the linear rate learning algorithm, we consider to first guess the labels for each instance so that the number of correct labels goes to infinity with the training sample size. Following the notation in Section 1.2.3, we let D be a distribution over $\mathcal{Y}$ with full support. For $((X_i, Y_i), Y_i')_{i \in \mathbb{N}} \sim (P \times D)^{\mathbb{N}}$, let $\boldsymbol{Z}_n := ((X_i, Y_i) : i \in [n], Y_i = Y_i')$ denote those in the first n examples whose labels are guessed correctly. Let $K_n := |\boldsymbol{Z}_n|$ denote the size of $\boldsymbol{Z}_n$. Then, conditional on K_n, $\boldsymbol{Z}_n$ follows the distribution $(P_D)^{K_n}$. Given training sample $S_n = ((X_i, Y_i))_{i=1}^n$, our learning algorithm $\mathbf{A}$ under bandit feedback applies $\mathcal{A}$ to the input sequence $\boldsymbol{Z}_n$. By the full supportedness of D, we can show that $K_n \to \infty$ almost surely. Then, by the universal consistency of $\mathcal{A}$, we can show that $\mathbb{E}[\mathrm{er}_{P_D}(\hat{h}_{\mathbf{A}}^{S_n})] \to 0$. To translate from $\mathbb{E}[\mathrm{er}_{P_D}(\hat{h}_{\mathbf{A}}^{S_n})]$ to $\mathbb{E}[\mathrm{er}_P(\hat{h}_{\mathbf{A}}^{S_n})]$, we can write $\mathbb{E}[\mathrm{er}_P(\hat{h}_{\mathbf{A}}^{S_n})]$ as an infinite sum over the index set $\mathcal{Y}$ by decomposing the error region based on the its intersection with $\mathcal{Z}_y := \mathcal{X} \times \{y\}$ for each $y \in \mathcal{Y}$. Then, we show that each term in the sum converges to zero by upper bounding it using $\mathbb{E}[\mathrm{er}_{P_D}(\hat{h}_{\mathbf{A}}^{S_n})]$ as D has full support. Finally, we can apply dominated converge to show that their sum converges to zero. See Section F for details.

1.3. Examples

First, we demonstrate that PAC learnability under bandit feedback implies universal learnability under bandit feedback at the Linear rate. Formally, we have the following proposition.

Proposition 7 (PAC Learnability $\implies$ Universal Learnability at $\frac{1}{n}$ rate) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass learning under bandit feedback framework. Further, assume that $\mathcal{Q}$ is PAC learnable under bandit feedback. Then, $\mathcal{Q}$ is universally learnable under bandit feedback at the rate $\frac{1}{n}$.*

Proof First, we claim that $\mathrm{ND}(\mathcal{C}) < \infty$. This is because if $\mathcal{Q}$ is PAC learnable under bandit feedback, then it is PAC learnable. In addition, if $\mathcal{Q}$ is PAC learnable, then $\mathrm{ND}(\mathcal{C}) < \infty$ (Daniely, Sabato, Ben-David, and Shalev-Shwartz, 2011). As a result, $\mathcal{C}$ does not have an infinite Natarajan Littlestone tree. Second, we claim that $\sup_{x \in \mathcal{X}} \{y \mid c \in \mathcal{C}, y = c(x)\} < \infty$. This is because if $\mathcal{Q}$ is PAC learnable under bandit feedback, then the effective label space should be bounded by Theorem 25. As a result, $\mathcal{C}$ does not have an infinite Progressive Littlestone tree. Therefore, based on the two facts mentioned combined with Theorem 3, we conclude that $\mathcal{Q}$ is universally learnable under bandit feedback at the rate $\frac{1}{n}$. This finishes the proof. ■

Next, we claim that PAC learnability under bandit feedback does not necessarily imply universal learnability under bandit feedback at the Exponential rate, and universal learnability under bandit feedback at the Exponential rate does not necessarily imply PAC learnability under bandit feedback. In fact, this claim is even provable in the binary setting where bandit feedback is equivalent to full supervision. In particular, the claim was proved by Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021, Example 2.2 and Example 2.3). Furthermore, in that work, they also provide an example (Example 2.1) that is both universally learnable at the rate e^{-n} and PAC learnable.

Proposition 8 (Universal Learnability at e^{-n} rate and PAC Learnability are not comparable) *There exists an instance of the multiclass learning under bandit feedback framework $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ such that it is PAC learnable under bandit feedback, but it is only universally learnable under bandit feedback at the optimal rate $\frac{1}{n}$. For the other direction, there exists an instance of the multiclass learning under bandit feedback framework $\mathcal{Q}' = (\mathcal{X}', \mathcal{Y}', \mathcal{C}')$ such that it is universally learnable under bandit feedback at the optimal rate e^{-n}, but it is not PAC learnable under bandit feedback.*

Subsequently, we give an instance of the multiclass learning under bandit feedback framework such that it is universally learnable under bandit feedback with an exponential rate, but it has infinite effective label space.

Proposition 9 (Universal Learnability at e^{-n} rate $\not\implies$ Finite Effective Label Space) *There exists an instance of the multiclass learning under bandit feedback framework $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ such that it is universally learnable under bandit feedback at the optimal rate at e^{-n}, but it has infinite effective label space.*

Proof Let $\mathcal{X} = \mathbb{R}$. Also, let $\mathcal{Y} = \mathbb{N}$. In addition, let $\mathcal{C} = \{c_y \mid c_y : \mathcal{X} \to \{y\}, y \in \mathcal{Y}\}$. First, we claim that $\sup_{x \in \mathcal{X}} \{y \mid c \in \mathcal{C}, y = c(x)\} = \infty$. This is because $\mathcal{Y} = \mathbb{N}$. Second, we claim that $\mathcal{C}$ does not have an infinite Littlestone tree. This is because once we fix a root $x \in \mathcal{X}$ of a Littlestone tree with right edge y_1, the class $c \mid c \in \mathcal{C}, c(x) = y_1$ has only one member, so the corresponding subtree cannot be infinite. Therefore, based on the two facts mentioned combined with Theorem 3,

we conclude that $\mathcal{Q}$ is universally learnable under bandit feedback at the optimal rate at e^{-n}, but has infinite effective label space. This finishes the proof. ■

Finally, we prove that having an infinite List Littlestone tree for every list size $\mathrm{L} \in \mathbb{N}$ does not imply having an infinite Progressive Littlestone tree.

Proposition 10 *There exists an instance of the multiclass learning under bandit feedback framework $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ such that it has an infinite List Littlestone tree for every list size $\mathrm{L} \in \mathbb{N}$, but does not have an infinite Progressive Littlestone tree.*

Proof Let $\mathcal{X} = \mathbb{N} \times \mathbb{R}$. Also, let $\mathcal{Y} = \mathbb{N} \cup \{\star\}$. Now, for every $n \in \mathbb{N}, n \geqslant 2$, we construct $\mathcal{C}_n$ consisting of functions from $\mathcal{X}$ to $\mathcal{Y}$ as follows: Let $\mathcal{T}_n$ be an infinite depth rooted perfect n-ary tree so that all of its nodes are labeled by distinct elements from $\{n\} \times \mathbb{R}$ and for every node in $\mathcal{T}$ all of the n outgoing edges are labeled by distinct elements form $\{1, 2, \ldots, n\}$. The definition of such a tree is similar to Definition 1.7 in the work of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). Based on that, construct $\mathcal{C}_n \subseteq \mathcal{Y}^{\mathcal{X}}$ as it contains only concepts consistent with all branches of $\mathcal{T}_n$ with each being $\star$ on every $x \in \mathcal{X}$ outside the corresponding branch. This finishes the construction. Let $\mathcal{C} = \cup_{i=2}^{\infty} \mathcal{C}_i$. Building on these, let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$. Indeed, $\mathcal{Q}$ has an infinite List Littlestone tree for every list size $\mathrm{L} \in \mathbb{N}$. This is because for every list size $\mathrm{L} \in \mathbb{N}$, the class $\mathcal{C}_{\mathrm{L}+1}$ can witness such an infinite tree based on our construction. Next, we claim that $\mathcal{Q}$ does not have an infinite Progressive Littlestone tree. We prove this by contradiction. Suppose we have an infinite progressive tree $\mathcal{T}$. This tree has a root $x = (i, c) \in \mathcal{X}$. Observe that if x is not a node label in $\mathcal{T}_i$, then we have a contradiction. This is because every $c \in \mathcal{C}$ should be $\star$ on x. Also, observe that if it is a node label in $\mathcal{T}$, then we still have a contradiction. This is because $|\{y \mid c \in \mathcal{C}, y = c(x)\}| = i$. Indeed, we have: $|\{y \mid c \in \mathcal{C}, y = c(x)\}| \geqslant i$ because x is a node label in $\mathcal{T}_i$. But we also have: $|\{y \mid c \in \mathcal{C}, y = c(x)\}| \leqslant i$ because x is not a node label on any tree other than $\mathcal{T}_i$. This finishes the proof. ■

1.4. Uniform Rates vs. Universal Rates

In this subsection, we further clarify the distinction between the PAC and universal learning frameworks. A good way to do so is to compare the definitions. Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass universal learning under bandit feedback framework. Also, let $\mathcal{R}$ be a rate function. Further, let $\mathcal{A}$ be the set of all learning algorithms.

First, we present the definition of uniform learnability within the realizable setting, which is connected to the realizable PAC learning framework. We say that $\mathcal{Q}$ is uniform learnable under bandit feedback at rate $\mathcal{R}$, if we have:

$$\exists_{\mathbf{A} \in \mathcal{A}} \ \exists\, C, c \in \mathbb{R}^{+} \ \ \forall\, \mathcal{D} \in \mathrm{RE}(\mathcal{C}) \ \ \forall n \in \mathbb{N}: \ \ \mathbb{E}_{\mathcal{S} \sim \mathcal{D}^n} \left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) \right] \leqslant C\mathcal{R}(cn).$$

Now, we recall the definition of universal learnability within the realizable setting, which is connected to the realizable universal learning framework. We say that $\mathcal{Q}$ is universally learnable under bandit feedback at rate $\mathcal{R}$, if we have:

$$\exists_{\mathbf{A} \in \mathcal{A}} \ \ \forall\, \mathcal{D} \in \mathrm{RE}(\mathcal{C}) \ \ \exists\, C, c \in \mathbb{R}^{+} \ \ \forall n \in \mathbb{N}: \ \ \mathbb{E}_{\mathcal{S} \sim \mathcal{D}^n} \left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) \right] \leqslant C\mathcal{R}(cn)$$

The definition of universal learnability can be seen as a simple rearrangement of the quantifiers in the definition of uniform learnability. Surprisingly, this seemingly minor modification significantly alters the problem's landscape (Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff, 2021). Notably, in the second definition, the constants $C, c \in \mathbb{R}^+$ are allowed to depend on the realizable distribution $\mathcal{D}$, whereas in the first definition, these constants must remain independent of $\mathcal{D}$.

1.5. Conclusion, Discussion, and Future Directions

In this manuscript, we extensively study the fundamental problem of multiclass learning with bandit feedback when the number of labels can be unbounded under the framework of *universal learning* (Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff, 2021) in the realizable setting. We have shown a first theoretical result in this context, stating a *trichotomy* of instance optimal learning curves. More specifically, the best achievable universal learning rate for any given concept class can only decay either at an *exponential*, a *linear*, or an *arbitrarily slow* rate. Furthermore, we exhibit appropriate combinatorial objects characterizing the mentioned cases, and we develop novel algorithms achieving instance optimal rates. Next, we provide a suggestion for future work.

- Exploring alternative forms of feedback beyond bandit feedback in multiclass online learning within the universal learning framework would be highly valuable. Examples of such feedback include apple tasting Helmbold, Littlestone, and Long (2000), dynamic pricing Cesa-Bianchi and Lugosi (2006), police and criminals Alon, Cesa-Bianchi, Dekel, and Koren (2015), matching pennies Lattimore and Szepesvári (2020), and feedback graphs Mannor and Shamir (2011). Can we provide a clean, unified theory? Notably, for instance, it is not hard to see that in the case of apple tasting feedback, we have an equivalence with the full supervision setting in universal learning framework.

1.6. Organization

The remainder of the paper is organized as follows. In Section A, we discuss a wider range of related works. In Section B, we formally set the notations and definitions. Subsequently, in Section C, we present our lower bounds. Following this, in Section D, we give the exponential upper bound. Then, in Section E, we prove the linear upper bound. Finally, in Section F, we present the universally consistent algorithm for our framework.

References

Alekh Agarwal, Sarah Bird, Markus Cozowicz, Luong Hoang, John Langford, Stephen Lee, Jiaji Li, Dan Melamed, Gal Oshri, Oswaldo Ribas, et al. Making contextual decisions with low technical debt. *arXiv preprint arXiv:1606.03966*, 2016.

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.

András Antos and Gábor Lugosi. Strong minimax lower bounds for learning. In *Proceedings of the ninth annual conference on Computational learning theory*, pages 303–309, 1996.

Dave Applebaum. Measure theory, by donald l. cohn. pp 373. swfr 74. 1993. isbn 0-8176-3003-1 3-7643-3003-1 (birkhäuser). *The Mathematical Gazette*, 79(484):222–223, 1995. doi: 10.2307/3620102.

Idan Attias, Steve Hanneke, Alkis Kalavasis, Amin Karbasi, and Grigoris Velegkas. Universal rates for regression: Separations between cut-off and absolute loss. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 359–405. PMLR, 2024.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002a.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002b.

Shai Ben-David and Nadav Eiron. Self-directed learning and its relation to the vc-dimension and to teacher-directed learning. *Mach. Learn.*, 33(1):87–104, 1998. doi: 10.1023/A:1007510732151.

Shai Ben-David, Nicolo Cesa-Bianchi, and Philip M Long. Characterizations of learnability for classes of {O,. . . , n}-valued functions. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 333–340, 1992.

Shai Ben-David, Nadav Eiron, and Eyal Kushilevitz. On self-directed learning. In Wolfgang Maass, editor, *Proceedings of the Eigth Annual Conference on Computational Learning Theory, COLT 1995, Santa Cruz, California, USA, July 5-8, 1995*, pages 136–143. ACM, 1995. doi: 10.1145/225298.225314.

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Learnability and the vapnik-chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989.

Olivier Bousquet, Steve Hanneke, Shay Moran, Ramon Van Handel, and Amir Yehudayoff. A theory of universal learning. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 532–541, 2021.

Olivier Bousquet, Steve Hanneke, Shay Moran, Jonathan Shafer, and Ilya Tolstikhin. Fine-grained distribution-dependent learning curves. *arXiv preprint arXiv:2208.14615*, 2022.

Nataly Brukhim, Elad Hazan, Shay Moran, Indraneel Mukherjee, and Robert E Schapire. Multiclass boosting and the cost of weak learning. *Advances in Neural Information Processing Systems*, 34: 3057–3067, 2021.

Nataly Brukhim, Daniel Carmon, Irit Dinur, Shay Moran, and Amir Yehudayoff. A characterization of multiclass learnability. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 943–955. IEEE, 2022.

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

David Cohn and Gerald Tesauro. Can neural networks do better than the vapnik-chervonenkis bounds? *Advances in Neural Information Processing Systems*, 3, 1990.

David Cohn and Gerald Tesauro. How tight are the vapnik-chervonenkis bounds? *Neural Computation*, 4(2):249–269, 1992.

Amit Daniely and Shai Shalev-Shwartz. Optimal learners for multiclass problems. In *Conference on Learning Theory*, pages 287–316. PMLR, 2014.

Amit Daniely, Sivan Sabato, Shai Ben-David, and Shai Shalev-Shwartz. Multiclass learnability and the erm principle. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 207–232. JMLR Workshop and Conference Proceedings, 2011.

Amit Daniely, Sivan Sabato, and Shai Shwartz. Multiclass learning approaches: A theoretical comparison with implications. *Advances in Neural Information Processing Systems*, 25, 2012.

Ofir David, Shay Moran, and Amir Yehudayoff. Supervised learning through the lens of compression. *Advances in Neural Information Processing Systems*, 29, 2016.

Pramith Devulapalli and Steve Hanneke. The dimension of self-directed learning. In Claire Vernade and Daniel Hsu, editors, *International Conference on Algorithmic Learning Theory, 25-28 February 2024, La Jolla, California, USA*, volume 237 of *Proceedings of Machine Learning Research*, pages 544–573. PMLR, 2024.

Liad Erez, Alon Cohen, Tomer Koren, Yishay Mansour, and Shay Moran. Fast rates for bandit pac multiclass classification. *arXiv preprint arXiv:2406.12406*, 2024a.

Liad Erez, Alon Cohen, Tomer Koren, Yishay Mansour, and Shay Moran. The real price of bandit information in multiclass classification. *arXiv preprint arXiv:2405.10027*, 2024b.

Dylan J Foster, Sham M Kakade, Jian Qian, and Alexander Rakhlin. The statistical complexity of interactive decision making. *arXiv preprint arXiv:2112.13487*, 2021.

David Gale and F. M. Stewart. Infinite games with perfect information. In *Contributions to the theory of games, vol. 2*, Annals of Mathematics Studies, no. 28, pages 245–266. Princeton University Press, Princeton, N.J., 1953.

Sally A. Goldman and Robert H. Sloan. The power of self-directed learning. *Mach. Learn.*, 14(1): 271–294, 1994. doi: 10.1023/A:1022605628675.

Sally A. Goldman, Ronald L. Rivest, and Robert E. Schapire. Learning binary relations and total orders. *SIAM J. Comput.*, 22(5):1006–1034, 1993. doi: 10.1137/0222062.

Steve Hanneke, Aryeh Kontorovich, Sivan Sabato, and Roi Weiss. Universal Bayes consistency in metric spaces. *The Annals of Statistics*, 49(4):2129 – 2150, 2021. doi: 10.1214/20-AOS2029.

Steve Hanneke, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Universal rates for interactive learning. *Advances in Neural Information Processing Systems*, 35:28657–28669, 2022.

Steve Hanneke, Shay Moran, and Qian Zhang. Universal rates for multiclass learning. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 5615–5681. PMLR, 12–15 Jul 2023.

Steve Hanneke, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Universal rates for active learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024a.

Steve Hanneke, Shay Moran, and Qian Zhang. Improved sample complexity for multiclass PAC learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b.

Steve Hanneke, Amirreza Shaeiri, and Hongao Wang. For universal multiclass online learning, bandit feedback and full supervision are equivalent. In *36th International Conference on Algorithmic Learning Theory*, 2024c.

D. Haussler, N. Littlestone, and M. Warmuth. Predicting $\{0, 1\}$-functions on randomly drawn points. *Information and Computation*, 115(2):248–292, 1994.

David Haussler. Decision theoretic generalizations of the pac model for neural net and other learning applications. *Information and computation*, 100(1):78–150, 1992.

David Haussler and Philip M Long. A generalization of sauer's lemma. *Journal of Combinatorial Theory, Series A*, 71(2):219–240, 1995.

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

Alkis Kalavasis, Grigoris Velegkas, and Amin Karbasi. Multiclass learnability beyond the pac framework: Universal rates and partial concept classes. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 20809–20822. Curran Associates, Inc., 2022.

Alexander S Kechris. Classical descriptive set theory. *Graduate Texts in Mathematics*, 156, 1995.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2:285–318, 1988.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. *Advances in Neural Information Processing Systems*, 24, 2011.

Shay Moran, Ohad Sharon, Iska Tsubari, and Sivan Yosebashvili. List online classification, 2023.

Balas K Natarajan. On learning sets and functions. *Machine Learning*, 4:67–97, 1989.

Balas K Natarajan and Prasad Tadepalli. Two new frameworks for learning. In *Machine Learning Proceedings 1988*, pages 402–415. Elsevier, 1988.

Herbert E. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535, 1952.

Benjamin Rubinstein, Peter Bartlett, and J Rubinstein. Shifting, one-inclusion mistake bounds and tight multiclass expected risk bounds. *Advances in Neural Information Processing Systems*, 19, 2006.

Dale Schuurmans. Characterizing rational versus exponential learning curves. *journal of computer and system sciences*, 55(1):140–160, 1997.

Ambuj Tewari and Susan A Murphy. From ads to interventions: Contextual bandits in mobile health. *Mobile health: sensors, analytic methods, and applications*, pages 495–517, 2017.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Vladimir Vapnik. On the uniform convergence of relative frequencies of events to their probabilities. In *Doklady Akademii Nauk USSR*, volume 181, pages 781–787, 1968.

Vladimir Vapnik. *Estimation of dependences based on empirical data*. Springer Science & Business Media, 2006.

Tom Viering and Marco Loog. The shape of learning curves: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Michael Woodroofe. A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806, 1979.

Contents

Appendix A. Related Work

Universal Learning. The notion of universal consistency has been a long-standing focus of research, with its roots in the statistics community. Nevertheless, the notion of universal learnability that we examine in the current manuscript was only recently introduced and popularized by the seminal work of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021). In fact, two works by Antos and Lugosi (1996) and Schuurmans (1997) made the first steps towards understating specific universal rates, where they explored special examples/cases. Since the work of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021), considerable efforts have been devoted to extending their result across various learning theoretic settings, including multiclass classification Kalavasis, Velegkas, and Karbasi (2022); Hanneke, Moran, and Zhang (2023), interactive learning Hanneke, Karbasi, Moran, and Velegkas (2022), real-valued regression Attias, Hanneke, Kalavasis, Karbasi, and Velegkas (2024), and active learning Hanneke, Karbasi, Moran, and Velegkas (2024a). Also, see Bousquet, Hanneke, Moran, Shafer, and Tolstikhin (2022). Notably, very recently, the work of Hanneke, Shaeiri, and Wang (2024c) studied multiclass online learning under bandit feedback within the framework of universal learning. Moreover, our exponential upper bound is built upon their online algorithm.

PAC Learning. The Probably Approximately Correct (PAC) learning framework, introduced by Valiant (1984), has been a foundational notion in statistical learning theory. The characterization of learnable classes in the binary PAC learning framework within the realizable setting was proved through the use of a combinatorial complexity parameter known as the VC dimension Blumer, Ehrenfeucht, Haussler, and Warmuth (1989); Vapnik (1968); Valiant (1984); Vapnik (2006). This result was later generalized to the agnostic setting by Haussler (1992). Since its inception, PAC learning has been extensively explored across various theoretical settings.

Bandit Feedback. The bandit setting plays a central role in statistical decision-making. The conceptual foundation of the bandit setting was first established by Thompson (1933). The field gained prominence through the influential work of Herbert Robbins, particularly his seminal paper Robbins (1952), which introduced the multi-armed bandit problem. In recent years, bandit scenarios have garnered significant attention, notably following the contributions of Auer, Cesa-Bianchi, and Fischer (2002a); Auer, Cesa-Bianchi, Freund, and Schapire (2002b). For a comprehensive literature review, see the recent work of Foster, Kakade, Qian, and Rakhlin (2021).

Multiclass Learning. A significant body of theoretical research has focused on multiclass classification across various frameworks, with contributions from Natarajan and Tadepalli (1988); Natarajan (1989); Ben-David, Cesa-Bianchi, and Long (1992); Haussler and Long (1995); Rubinstein, Bartlett, and Rubinstein (2006); Daniely, Sabato, Ben-David, and Shalev-Shwartz (2011); Daniely, Sabato, and Shwartz (2012); Daniely and Shalev-Shwartz (2014); Brukhim, Hazan, Moran, Mukherjee, and Schapire (2021). However, the combinatorial characterization of multiclass classification within Valiant's PAC learning framework, particularly when the number of labels can be unbounded, remained unresolved until recently, even in the realizable setting. This open question was addressed in the seminal work of Brukhim, Carmon, Dinur, Moran, and Yehudayoff (2022). Moreover, the same dimension also characterizes the agnostic version of the problem David, Moran, and Yehudayoff (2016).
Notably, several factors motivate the study of multiclass classification with unbounded label spaces. First, in multiclass settings, it is desirable for guarantees to remain independent of the number of

labels, even when finite. Second, mathematical frameworks involving infinity often provide clearer and more elegant insights. Finally, from a practical perspective, many essential machine learning tasks require classification into extremely large label spaces, such as in image object recognition.

Appendix B. Notations, Definitions, and Preliminaries

First, we set our basic notation in Section B.1. Subsequently, we define multiclass learning under bandit feedback framework in Section B.2. Finally, we give definitions of our combinatorial complexity parameters in Section B.3.

B.1. Notations

In this subsection, we present the basic notations that we use throughout our paper. Let $\mathbb{N}$ and $\mathbb{R}$ stand for the set of natural numbers and real numbers, accordingly. Also, for a given $n \in \mathbb{N}$, we use $[n]$ to denote $\{1, 2, \ldots, n\}$. Next, let $n \in \mathbb{N}$, for any sequence of size n or n-tuple x, and any $i \in \mathbb{N}$ such that $1 \leqslant i \leqslant n$, let us use x_i to denote the i-th element in x. To increase the readability of our manuscript, we use "," to separate indices of elements when we have more than one index; for instance, let x be a sequence of size 5 of 2-tuples, we denote by $x_{5,1}$ the first element of the 5-th element of x. Also, let $m, n \in \mathbb{N}$ such that $m \leqslant n$, we write $\big((x_i, y_i)\big)_{i=m}^{n}$ to denote $\big((x_m, y_m), (x_{m+1}, y_{m+1}), \ldots, (x_n, y_n)\big)$. Afterward, we denote by $A \times B$ the Cartesian product of two arbitrarily set A and B. In addition, for any set A and any $n \in \mathbb{N}$, we let A^n indicate n times the Cartesian product of A with itself. Note that for any set A, we define $A^0 := \{\varnothing\}$. Also, given a set A, we denote by $A^\star$ the set of all finite sequences of members of A; more formally, $A^\star := \bigcup_{T=0}^{\infty} A^T$. Then, for the arbitrary set X and Y, we use Y^X to denote the space of all functions from X to Y. Going further, let e and $\log(\cdot)$ stand for Euler's number and Logarithm function in base 2, respectively. Finally, we use $\mathcal{O}(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, $\omega(\cdot)$, and $\Theta(\cdot)$ as standard notations of them in theoretical computer science.

B.2. Multiclass Universal Learning under bandit feedback Framework

In this subsection, we formally and rigorously define multiclass learning under bandit feedback framework. We may restate some definitions that we have presented in Section 1.1 for the sake of completeness.

B.2.1. PROBLEM SETUP

Fix a non-empty Polish space $\mathcal{X}$ as an instance space Definition 11. For example, the instance space $\mathcal{X}$ can be a Euclidean space or any countable set. Also, fix a non-empty and countable set $\mathcal{Y}$ as a label space. Moreover, for the power set σ-algebra $\sigma(\mathcal{Y})$ over the label space $\mathcal{Y}$, define $\Pi(\mathcal{Y})$ as the set of all probability measures on $(\mathcal{Y}, \sigma(\mathcal{Y}))$. A pair $z = (x, y) \in \mathcal{X} \times \mathcal{Y}$ is called an example, and $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ denotes the space of examples. A concept is a function from $\mathcal{X}$ to $\mathcal{Y}$. With this in mind, fix a non-empty set of concepts $\mathcal{C}$ that satisfies a minimal measurability assumption stated in Definition 12 as a concept class. In particular, a 3-tuple $\mathcal{Q} = \big(\mathcal{X}, \mathcal{Y}, \mathcal{C}\big)$ presents an instance of the multiclass learning under bandit feedback framework.

Definition 11 (Polish Spaces) *A Polish space is a separable topological space that admits a complete metric.*

Definition 12 (Measurable Concept Class) *Fix Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. A set $\mathcal{C} \subseteq \mathcal{Y}^{\mathcal{X}}$ is said to be measurable if there is a Polish space Θ and a Borel-measurable map $\mathsf{h} : \Theta \times \mathcal{X} \to \mathcal{Y}$ such that $\mathcal{C} = \{\mathsf{h}(\theta, \cdot) : \theta \in \Theta\}$.*

The Borel isomorphism theorem (Applebaum, 1995, Theorem 8.3.6) implies that we would obtain an identical definition if we required only that Θ is a Borel subset of a Polish space.

Notably, the assumption stated above holds for almost any $\mathcal{C}$ encountered in practical applications. Furthermore, it is important to highlight that this definition represents the standard measurability assumption commonly used in empirical process theory, where it is typically referred to as the image-admissible Suslin property.

Next, we define non-degenerate instances of the multiclass universal learning under bandit feedback framework. to remove trivial cases.

Definition 13 (Non-degenerate Instances) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass universal learning under bandit feedback framework. We say $\mathcal{Q}$ is a non-degenerate instance of the multiclass universal learning under bandit feedback framework if there exists $x_1, x_2 \in \mathcal{X}$ and $c_1, c_2 \in \mathcal{C}$ such that $c_1(x_1) = c_2(x_1)$ and $c_1(x_2) \neq c_2(x_2)$.*

B.2.2. MULTICLASS LEARNING UNDER BANDIT FEEDBACK GAME

A data distribution $\mathcal{D}$ is a probability measure on $\mathcal{Z}$. Initially, nature chooses a data distribution $\mathcal{D}$. Fix $n \in \mathbb{N}$ as the number of samples. Then, we have a n-rounded sequential game between the learner and nature. Moreover, at each round $t \in [n]$:

1. Nature samples an instance (x_t, y_t) from $\mathcal{D}$ and reveals x_t to the learner.
2. The learner predicts a label $\hat{y}_t$ from $\mathcal{Y}$.
3. Nature reveals the feedback $f_t = \mathbb{1}\{y_t \neq \hat{y}_t\}$ to the learner.

Finally, perhaps based on $\{(x_1, \hat{y}_1, f_1), (x_2, \hat{y}_2, f_2), \ldots, (x_n, \hat{y}_n, f_n)\}$, the learner outputs a universally measurable function from $\mathcal{X}$ to $\mathcal{Y}$.

B.2.3. LEARNING ALGORITHMS

Let Σ be defined as $\Sigma := \mathcal{Y} \times \{0, 1\}$. A learning algorithm $\mathbf{A}$ is a 2-tuple of mappings. In particular, $\mathbf{A}_1$ is a mapping from $(\mathcal{X} \times \Sigma)^\star \times \mathcal{X}$ to a probability measure $\nu \in \Pi(\mathcal{Y})$. In addition, $\mathbf{A}_2$ is also a mapping from $(\mathcal{X} \times \Sigma)^\star \times \mathcal{X}$ to a probability measure $\mu \in \Pi(\mathcal{Y})$. We define two separate mappings,

one related to the game and one related to the final result, because we believe that this approach is more convenient to understand.

Now, let $\mathbf{A}$ be a learning algorithm. For any $\mathfrak{u} \in \big(\mathcal{X} \times \Sigma\big)^\star$ and any $x \in \mathcal{X}$, let us write $\mathbf{A}_1(x; \mathfrak{u})$ to denote $\mathbf{A}_1\big((\mathfrak{u}, x)\big)$. In addition, for every finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, we denote by $\mathbf{B}_{\mathbf{A}}(\mathcal{S})$ the random variable taking values in $(\mathcal{X} \times \Sigma)^n$ representing the first part of the input of $\mathbf{A}_1$ in round $n+1$ when the choices of nature are obtained according to $\mathcal{S}$, which is recursively defined as follows:

$$\mathbf{B}_{\mathbf{A}}(\mathcal{S})_{i,1} := \mathcal{S}_{i,1}, \quad 1 \leqslant i \leqslant n$$

$$\mathbf{B}_{\mathbf{A}}(\mathcal{S})_{i,2} := \begin{cases} y \sim \mathbf{A}_1\big(\mathcal{S}_{1,1}; \{\varnothing\}\big), & i = 1 \\ y \sim \mathbf{A}_1\big(\mathcal{S}_{i,1}; \big[\big(\mathbf{B}_{\mathbf{A}}(\mathcal{S})_{j,1}, \mathbf{B}_{\mathbf{A}}(\mathcal{S})_{j,2}, \mathbf{B}_{\mathbf{A}}(\mathcal{S})_{j,3}\big)\big]_{j=1}^{i-1}\big), & 1 < i \leqslant n \end{cases}$$

$$\mathbf{B}_{\mathbf{A}}(\mathcal{S})_{i,3} := \mathbb{1}\big\{\mathbf{B}_{\mathbf{A}}(\mathcal{S})_{i,2} \neq \mathcal{S}_{i,2}\big\}, \quad 1 \leqslant i \leqslant n$$

Building upon that, also for every finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, we denote by $\hat{h}_{\mathbf{A}}^{\mathcal{S}}$ the random function taking values in $\mathcal{Y}^{\mathcal{X}}$ such that for every $x \in \mathcal{X}$, we have: $\hat{h}_{\mathbf{A}}^{\mathcal{S}}(x) = \hat{y}$, where $\hat{y} \sim \mathbf{A}_2((\mathbf{B}_{\mathbf{A}}(\mathcal{S}), x))$. Note that whenever it is clear from the context, we may simply write $\hat{h}$ instead of $\hat{h}_{\mathbf{A}}^{\mathcal{S}}$.

Let $\mathbf{A}$ be a learning algorithm. Notably, in this paper, we focus on deterministic $\mathbf{A}_2$. However, our results remain valid when randomized $\mathbf{A}_2$ are allowed as well. In particular, all lower bounds we prove also hold for randomized $\mathbf{A}_2$.

Finally, we have the definition of universally measurable functions. In this paper, we only work with universally measurable functions.

Definition 14 (Universally Measurable Functions) *Let $\mathfrak{F}$ be the Borel σ-field on some Polish space $\mathcal{X}$ and let μ be a probability measure. We denote by $\mathfrak{F}_\mu$ the completion of $\mathfrak{F}$ under μ. In particular, the collections of all subsets of $\mathcal{X}$ that differ from a Borel set on a set of zero measure. A set $\mathcal{X}' \subseteq \mathcal{X}$ is called universally measurable if $\mathcal{X}' \in \mathfrak{F}_\mu$ for every probability measure μ. Moreover, a function $f : \mathcal{X} \to \mathcal{Y}$ is called universally measurable if $f^{-1}(\mathcal{Y}')$ is universally measurable, for any universally measurable subset $\mathcal{Y}'$ of $\mathcal{Y}$.*

B.2.4. ERROR RATE

For a data distribution $\mathcal{D}$ over $\mathcal{Z}$ and a concept $c : X \to \mathcal{Y}$, define the error rate of a concept c with respect to the data distribution $\mathcal{D}$ denoted by $\mathrm{er}_{\mathcal{D}}(c)$ as follows: $\mathrm{er}_{\mathcal{D}}(c) := \mathcal{D}(\{(x, y) \in \mathcal{Z} : c(x) \neq y\}) = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[c(X) \neq Y]$.

In particular, for a data distribution $\mathcal{D}$, a learning algorithm $\mathbf{A}$, and a finite sequence of examples $\mathcal{S} \in \mathcal{Z}^n = (\mathcal{X} \times \mathcal{Y})^n$ of size n for some $n \in \mathbb{N}$, the error rate of $\hat{h}_{\mathbf{A}}^{\mathcal{S}}$ with respect to the data distribution $\mathcal{D}$ is $\mathrm{er}_{\mathcal{D}}(\hat{h}_{\mathbf{A}}^{\mathcal{S}}) = \mathcal{D}(\{(x, y) \in \mathcal{Z} : \hat{h}_{\mathbf{A}}^{\mathcal{S}}(x) \neq y\})$.

Notably, the 0-1 loss function is recognized as a standard choice for assessing the performance of learning algorithms in multiclass learning problems within the statistical learning theory community.

B.2.5. REALIZABLE DISTRIBUTION

Based on the previous subsubsection, we say that a data distribution $\mathcal{D}$ is realizable by a concept class $\mathcal{C}$ if $\inf_{c\in\mathcal{C}} \ \mathrm{er}_{\mathcal{D}}(c) = 0$. Furthermore, we denote by $\mathrm{RE}(\mathcal{C})$ the set of all realizable data distributions. Notably, the realizabilty assumption is a standard assumption in the statistical learning theory literature.

Besides realizable distributions, we also introduce the notion of realizable sequences which is convenient to use. A finite sequence $((x_1, y_1), \ldots, (x_n, y_n)) \in \mathcal{Z}^n$ with $n \in \mathbb{N}$ is called $\mathcal{C}$-realizable if there exists some $h \in \mathcal{C}$ such that $h(x_i) = y_i$ for all $i \in [n]$. An infinite sequence $((x_1, y_1), (x_2, y_2), \ldots) \in \mathcal{Z}^{\mathbb{N}}$ is called $\mathcal{C}$-realizable if for any $t \in \mathbb{N}$, there exists some $h \in \mathcal{C}$ such that $h(x_i) = y_i$ for all $i \in [t]$. Note that we do not require the existence of some $h \in \mathcal{C}$ with $h(x_i) = y_i$ for all $i \in \mathbb{N}$.

B.2.6. UNIVERSAL LEARNABILITY

Informally, the objective of universal learnability in multiclass learning under bandit feedback framework is to design a learning algorithm $\mathbf{A}$ such that for every realizable distribution $\mathcal{D}$ the expected error rate $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right]$ decays as rapidly as possible as a function of the sample size n. We start with the following simple definition.

A function $\mathcal{R} : \mathbb{N} \to (0, 1]$ is a rate function if $\lim_{n\to\infty} \mathcal{R}(n) = 0$. Moreover, we use the subscript of n to emphasize that we have a rate function.

Now, based on the simple definition above, we formalize the objective in multiclass universal learning under bandit feedback framework.

Definition 15 *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass learning under bandit feedback framework. Also, let $\mathcal{R}$ be a rate function. Then, we say that:*

- *$\mathcal{Q}$ is universally learnable under bandit feedback at rate $\mathcal{R}$, if there exists a learning algorithm $\mathbf{A}$ such that for every realizable distribution $\mathcal{D}$, there exist $C, c \in \mathbb{R}^+$ such that for every $n \in \mathbb{N}$, we have: $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \leqslant C\mathcal{R}(cn)$.*
- *$\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}$, if for all learning algorithms $\mathbf{A}$, there exists a realizable distribution $\mathcal{D}$ and $C, c \in \mathbb{R}^+$ such that for infinitely many $n \in \mathbb{N}$, we have: $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \geqslant C\mathcal{R}(cn)$.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback at optimal rate $\mathcal{R}$, if $\mathcal{Q}$ is universally learnable under bandit feedback at rate $\mathcal{R}$ and $\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}$.*
- *$\mathcal{Q}$ is universally learnable under bandit feedback, if there exists a learning algorithm $\mathbf{A}$ such that for every realizable distribution $\mathcal{D}$, we have: $\lim_{n\to\infty} \mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] = 0$.*
- *$\mathcal{Q}$ requires arbitrarily slow rates under bandit feedback, if for every rate function $\mathcal{R}'$, we know that $\mathcal{Q}$ is not universally learnable under bandit feedback at rate faster than $\mathcal{R}'$.*

B.3. Combinatorial Complexity Parameters

In this subsection, we preset the definitions of the main combinatorial complexity parameters in our paper.

Definition 16 (Perfect Rooted L-ary Trees) *Let $\mathrm{L} \in \mathbb{N}, \mathrm{L} \geqslant 2$. A perfect rooted L-ary tree $\mathcal{T}$ is a rooted tree, each of whose internal nodes has exactly L children and all leaves have the same depth.*

Definition 17 (L-ary $(\mathscr{X}, \mathscr{Y})$-valued Trees) *Let $\mathrm{L} \in \mathbb{N}$. Also, let $\mathscr{X}, \mathscr{Y}$ be any non-empty sets. An L-ary $(\mathscr{X}, \mathscr{Y})$-valued tree $\mathcal{T}$ is a perfect rooted L-ary tree, each of whose nodes are labeled by an element of $\mathscr{X}$, and each of whose edges are labeled by an element of $\mathscr{Y}$. Moreover, for any L-ary $(\mathscr{X}, \mathscr{Y})$-valued tree, a root-to-leaf path can be identified by a sequence of pairs $\mathfrak{s} \in (\mathscr{X} \times \mathscr{Y})^\star$.*

Definition 18 ($(\mathrm{L}+1)$-Littlestone Tree and Multiclass Littlestone Tree) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass universal learning under bandit feedback framework. Let L in $\mathbb{N}$. We say that an $(\mathrm{L}+1)$-ary $(\mathcal{X}, \mathcal{Y})$-valued tree $\mathcal{T}$ is an $(\mathrm{L}+1)$-Littlestone tree for $\mathcal{C}$ if, (1) all children of every node have distinct labels, and (2) $\mathcal{T}$ is shattered by $\mathcal{C}$: for every finite root-to-leaf path in $\mathcal{T}$, identified by $\mathfrak{s} \in (\mathcal{X} \times \mathcal{Y})^\star$, there exists a concept $c \in \mathcal{C}$ such that for every $i \in [|\mathfrak{s}|]$, we have $\mathfrak{s}_{i,2} = c(\mathfrak{s}_{i,1})$.*

We say $\mathcal{C}$ has an infinite $(\mathrm{L}+1)$-Littlestone tree if there exists an $(\mathrm{L}+1)$-Littlestone tree for $\mathcal{C}$ of depth ∞.

A 2-Littlestone tree is also referred to as a multiclass Littlestone tree.

Definition 19 (Perfect Rooted Progressive Trees) *A perfect rooted progressive tree $\mathcal{T}$ is a rooted tree, each of whose internal nodes at depth $i \in \mathbb{N} \cup \{0\}$ has exactly $i+2$ children and all leaves have the same depth.*

Definition 20 (Progressive $(\mathscr{X}, \mathscr{Y})$-valued Trees) *Let $\mathscr{X}, \mathscr{Y}$ be any non-empty sets. A progressive $(\mathscr{X}, \mathscr{Y})$-valued tree $\mathcal{T}$ is a perfect rooted progressive tree, each of whose nodes are labeled by an element of $\mathscr{X}$, and each of whose edges are labeled by an element of $\mathscr{Y}$. Moreover, for any progressive $(\mathscr{X}, \mathscr{Y})$-valued tree, a root-to-leaf path of length can be identified by a sequence of pairs $\mathfrak{s} \in (\mathscr{X} \times \mathscr{Y})^\star$.*

Definition 21 (Progressive Littlestone Tree) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of the multiclass universal learning under bandit feedback framework. A progressive $(\mathcal{X}, \mathcal{Y})$-valued tree $\mathcal{T}$ is called* progressive Littlestone tree *for $\mathcal{C}$, if (1) all children of every node have distinct labels, and (2) $\mathcal{T}$ is shattered by $\mathcal{C}$: for every finite root-to-leaf path in $\mathcal{T}$, identified by $\mathfrak{s} \in (\mathcal{X} \times \mathcal{Y})^\star$, there exists a concept $c \in \mathcal{C}$ such that for every $i \in [|\mathfrak{s}|]$, we have $\mathfrak{s}_{i,2} = c(\mathfrak{s}_{i,1})$.*

We say $\mathcal{C}$ has an infinite progressive Littlestone tree if there exists a progressive Littlestone tree for $\mathcal{C}$ of depth ∞.

Definition 22 (Natarajan Shattered Natarajan (1989)) *Let $\mathcal{C} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a concept class. Let $S \subseteq \mathcal{X}$ be a set of instances. We say that S is* NT-shattered *by $\mathcal{C}$, if there exists $f, g : S \to \mathcal{Y}$, where $\forall_{x \in \mathcal{S}}\ f(x) \neq g(x)$ such that for every $T \subseteq S$ there exists $c \in \mathcal{C}$ such that:*

$$\forall_{x \in T}\ c(x) = f(x) \text{ and } \forall_{x \in S-T}\ c(x) = g(x)$$

Definition 23 (Natarajan Dimension) *Let $\mathcal{C} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a concept class. The* Natarajan dimension *of $\mathcal{C}$, denoted by $\mathrm{NT}(\mathcal{C})$, is defined as a $\sup_{d \in \mathbb{N}}$ such that there exists a set of instances $\mathcal{S} \subseteq \mathcal{X}$ of size d that is NT-shattered by $\mathcal{C}$. Also, if $\mathcal{C} = \{\varnothing\}$, we have: $\mathrm{NT}(\mathcal{C}) = 0$.*

Definition 24 (Natarajan Littlestone Tree, Kalavasis, Velegkas, and Karbasi 2022) *Let $\mathcal{C} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a concept class. A* Natarajan Littlestone (NL) tree *of depth $d \leqslant \infty$ for $\mathcal{C}$ consists of a tree*

$$\bigcup_{0 \leqslant \ell < d} \left\{ x_u \in \mathcal{X}^{\ell+1}, u \in \{0,1\} \times \{0,1\}^2 \times \ldots \times \{0,1\}^{\ell} \right\}$$

and two colorings $s^{(0)}, s^{(1)}$ mapping each position $u^i \in u$ for any node with pattern $u \in \{0,1\} \times \ldots \times \{0,1\}^{\ell}$ for $i \in \{0,1,\ldots,\ell\}$ and $\ell \in \{0,1,\ldots,d-1\}$ of the tree to some color $\mathcal{Y}$ such that for every finite level $n < d$, the subtree $T_n = \cup_{0 \leqslant \ell \leqslant n} \{x_u = (x_u^0, \ldots, x_u^{\ell}) : u \in \{0,1\} \times \{0,1\}^2 \times \ldots \times \{0,1\}^{\ell}\}$ satisfies the following:

1. *At any point $x_u^i \in x_u \in T_n$, it holds $s^{(0)}(x_u^i) \neq s^{(1)}(x_u^i)$ and*
2. *for any path $\boldsymbol{y} \in \{0,1\} \times \ldots \times \{0,1\}^{n+1}$, there exists a concept $c \in \mathcal{C}$ so that $c(x^i_{\boldsymbol{y}_{\leqslant \ell}}) = s^{(0)}(x^i_{\boldsymbol{y}_{\leqslant \ell}})$ if $y^i_{\ell+1} = 1$ and $c(x^i_{\boldsymbol{y}_{\leqslant \ell}}) = s^{(1)}(x^i_{\boldsymbol{y}_{\leqslant \ell}})$ otherwise, for all $0 \leqslant i \leqslant \ell$ and $0 \leqslant \ell \leqslant n$, where*
$$\boldsymbol{y}_{\leqslant \ell} = (y_1^0, (y_2^0, y_2^1), \ldots, (y_\ell^0, \ldots, y_\ell^{\ell-1})), x_{\boldsymbol{y}_{\leqslant \ell}} = (x^0_{\boldsymbol{y}_{\leqslant \ell}}, \ldots, x^{\ell}_{\boldsymbol{y}_{\leqslant \ell}}).$$

We say that $\mathcal{C}$ has an infinite NL tree if it has a NL tree of depth $d = \infty$.

Appendix C. Lower Bounds

We begin by proving that the finiteness of the effective label space is necessary for PAC learnability with bandit feedback. Formally, we have the following theorem.

Theorem 25 (PAC Learnability $\implies$ Finite Effective Label Space) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. Then, if $\mathcal{Q}$ is PAC learnable with bandit feedback, it has finite effective label space.*

Proof We prove this by contradiction. Suppose that there exists a learning algorithm $\mathbf{A}$ and a sample size $n \in \mathbb{N}$ such that for every realizable distribution $\mathcal{D}$, we have: $\mathbb{P}_{\mathcal{S} \sim \mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) \leqslant \frac{1}{2}\right] \geqslant \frac{1}{2}$ and $\sup_{x \in \mathcal{X}} \{y \mid c \in \mathcal{C}, y = c(x)\} = \infty$. Now, take $x \in \mathcal{X}$ such that we have: $\{y \mid c \in \mathcal{C}, y = c(x)\} \geqslant 2n+1$. Indeed, we can do so because $\sup_{x \in \mathcal{X}} \{y \mid c \in \mathcal{C}, y = c(x)\} = \infty$. Next, take $\mathcal{B}$ such that

$\mathcal{B} \subseteq \{y \mid c \in \mathcal{C}, y = c(x)\}$ and $|\mathcal{B}| = 2n + 1$. Subsequently, define P as the uniform distribution on $\mathcal{B}$. Note that the probability of guessing the correct label having n trials is at most $n/(2n+1) < 1/2$. Therefore, there exists a hard realizable distribution $\mathcal{D}$ such that $\mathbb{P}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) > \frac{1}{2}\right] > \frac{1}{2}$. This is a contradiction. As a result, if $\mathcal{Q}$ is PAC learnable with bandit feedback, it has finite effective label space. This finishes the proof. ■

Next, we draw on two theorems from the work of Hanneke, Moran, and Zhang (2023) on universal learnability under full supervision. Indeed, any lower bound established for multiclass universal learning under full supervision should apply to the bandit setting as well. Formally, we have the following theorems.

Theorem 26 (Faster than Exponential is not Possible) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. Then, $\mathcal{Q}$ is not universal learnable under bandit feedback at rate faster than exponential: $\mathcal{R}(n) = \mathrm{e}^{-n}$.*

Proof Fix a learning algorithm $\mathbf{A}$. Then, based on the work of Hanneke, Moran, and Zhang (2023), there exists a realizable distribution $\mathcal{P}_{XY}$ such that $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \geqslant 2^{-n-2}$ for infinitely many $n \in \mathbb{N}$, even if we have full supervision feedback. This finishes the proof. ■

Theorem 27 (Slower than Exponential is not Faster than Linear) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. Assume that $\mathcal{C}$ has an infinite multiclass Littlestone tree. Then, $\mathcal{Q}$ is not universal learnable under bandit feedback at rate faster than exponential: $\mathcal{R}(n) = \frac{1}{n}$.*

Proof Fix a learning algorithm $\mathbf{A}$. Then, based on the work of Hanneke, Moran, and Zhang (2023), there exists a realizable distribution $\mathcal{P}_{XY}$ such that $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \geqslant \frac{1}{32n}$ for infinitely many $n \in \mathbb{N}$, even if we have full supervision feedback. This finishes the proof. ■

Finally, we present the following lower bound that serves as the novel contribution of the current manuscript. Formally, we have the following theorems.

Theorem 28 (Slower than Linear is Arbitrarily Slow) *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. If $\mathcal{C}$ has either an infinite Natarajan Littlestone tree or an infinite progressive Littlestone tree, then $\mathcal{Q}$ requires arbitrarily slow rates.*

Proof Any multiclass learning algorithm $\mathbf{A}$ under the bandit feedback framework naturally defines a multiclass learning algorithm $\widetilde{\mathbf{A}}$ under the full-information feedback framework inductively as follows. For the initial step, we define the output multiclass classifier $\hat{h}_0 : \mathcal{X} \to \mathcal{Y}$ of $\widetilde{\mathbf{A}}$ to be $\hat{h}_0 = \widetilde{\mathbf{A}}(\varnothing) := \mathbf{A}(\varnothing)$. Suppose that for some $n \in \{0\} \cup \mathbb{N}$ and any sequence $s_n = ((x_1, y_1), \ldots, (x_n, y_n)) \in \mathcal{Z}^n$, we have defined the output classifiers $\hat{h}_t = \widetilde{\mathbf{A}}(s_t)$ of $\widetilde{\mathbf{A}}$ for all $0 \leqslant t \leqslant n$, where $s_t := ((x_1, y_1), \ldots, (x_t, y_t))$. Given a new instance $(x_{n+1}, y_{n+1}) \in \mathcal{Z}$, we define

$\hat{y}_{t+1} := \hat{h}_t(x_{t+1})$ for all $0 \leqslant t \leqslant n$, $s_{n+1} = ((x_t, y_t))_{t=1}^{n+1}$, $\hat{s}_{n+1} := ((x_t, \mathbb{1}\{y_t \neq \hat{y}_t\}, \hat{y}_t))_{t=1}^{n+1}$, and $\hat{h}_{n+1} = \widetilde{\mathbf{A}}(s_{n+1}) := \mathbf{A}(\hat{s}_{n+1})$. In this way, we have defined $\widetilde{\mathbf{A}}$ so that its output classifier is the same as that of $\mathbf{A}$ under the same adversary sequence. Thus, if $\mathcal{Q}$ requires arbitrarily slow rates under the full-information feedback framework, it will also require arbitrarily slow rates under the bandit feedback framework. Note that an infinite Natarajan Littlestone tree is an infinite DSL tree (Hanneke et al., 2023, Definition 7), and if $\mathcal{C}$ has an infinite DSL tree, $\mathcal{Q}$ will require arbitrarily slow rates under the full-information feedback framework (Hanneke et al., 2023, Theorem 15). It follows that $\mathcal{Q}$ requires arbitrarily slow rates under the bandit feedback framework if $\mathcal{C}$ has an infinite Natarajan Littlestone tree.

Now, suppose that $\mathcal{C}$ has an infinite progressive Littlestone tree. For any $k \in \{0\} \cup \mathbb{N}$, define $\mathcal{I}_k := [2] \times [3] \times \cdots \times [k+1]$ with the convention that $\mathcal{I}_0 = \varnothing$. Define $\mathcal{I} := \cup_{k=0}^{\infty} \mathcal{I}_k$. Then, we can represent the infinite progressive Littlestone tree of $\mathcal{Q}$ as

$$\mathcal{T} = \{x_{\boldsymbol{u}} : \boldsymbol{u} \in \mathcal{I}, x_{\boldsymbol{u}} \in \mathcal{X}\} \cup \{y_{\boldsymbol{u}} : \boldsymbol{u} \in \cup_{k=1}^{\infty} \mathcal{I}_k, y_{\boldsymbol{u}} \in \mathcal{Y}\}.$$

such that for any $\boldsymbol{u} \in \mathcal{I}_k$ with $k \geqslant 0$, $|\{x_{(\boldsymbol{u},i)} : i \in [k+2]\}| = k+2 = |\{y_{(\boldsymbol{u},i)} : i \in [k+2]\}|$. For any $k \geqslant 0$, sample $U_k \sim \mathrm{Unif}([k+1])$ and define the infinite random sequence $\boldsymbol{U} := (U_1, U_2, U_3, \dots)$. For any $k \in \mathbb{N}$, we use $\boldsymbol{U}_{\leqslant k}$ or $\boldsymbol{U}_{<(k+1)}$ to represent $(U_1, \dots, U_k)$. For $k = 0$, we define $\boldsymbol{U}_{\leqslant 0}$ and $\boldsymbol{U}_{<1}$ to be $\varnothing$. Let $\mathcal{R}$ be and arbitrary rate function. Let $(p_i)_{i\in\mathbb{N}}$, $(n_i)_{i\in\mathbb{N}}$, and $(k_i)_{i\in\mathbb{N}}$ be the sequences defined in Lemma 29 with $m = 4$. Let D denote the probability measure on $\mathbb{N}$ such that $D(\{i\}) = p_i$ for all $i \in \mathbb{N}$. For any $\boldsymbol{u} \in \mathcal{I}_\infty := [2] \times [3] \times [4] \times \cdots$, we let $P_{\boldsymbol{u}}$ denote the probability measure on $\mathcal{X} \times \mathcal{Y}$ such that $P_{\boldsymbol{u}}(\{x_{\boldsymbol{u}_{<k}}, y_{\boldsymbol{u}_{\leqslant k}}\}) = p_k$ for any $k \in \mathbb{N}$. Since $\mathcal{T}$ is shattered by $\mathcal{C}$, there exists $h_k \in \mathcal{C}$ for each $k \in \mathbb{N}$ such that $h_k(x_{\boldsymbol{u}_{<i}}) = y_{\boldsymbol{u}_{\leqslant i}}$ for all $i \in [k]$. It follows that

$$\inf_{h\in\mathcal{C}} \mathrm{er}_{P_{\boldsymbol{u}}}(h) \leqslant \mathrm{er}_{P_{\boldsymbol{u}}}(h_k) \leqslant \sum_{i=k+1}^{\infty} p_i.$$

Since $\sum_{i\in\mathbb{N}} p_i = 1$, we have $\inf_{h\in\mathcal{C}} \mathrm{er}_{P_{\boldsymbol{u}}}(h) = 0$, i.e., $P_{\boldsymbol{u}}$ is $\mathcal{C}$-realizable.

Sample $(K, K_1, \dots, K_n) \sim D^{n+1}$ with $n \in \mathbb{N}$ and let $X = x_{\boldsymbol{U}_{<K}}$, $Y = y_{\boldsymbol{U}_{\leqslant K}}$, $X_i = x_{\boldsymbol{U}_{<K_i}}$, and $Y_i = y_{\boldsymbol{U}_{\leqslant K_i}}$ for any $i \in [n]$. Then, conditional on $\boldsymbol{U}$, we have $((X, Y), (X_1, Y_1), \dots, (X_n, Y_n)) \sim (P_{\boldsymbol{U}})^{n+1}$. Define $S_i := ((X_j, Y_j))_{j=1}^{i}$ for $i \in [n]$, $S_0 := \varnothing$, $\hat{h}_i := \widetilde{\mathbf{A}}(S_i)$ for $0 \leqslant i \leqslant n$, $\hat{Y}_i := \hat{h}_{i-1}(X_i)$ for $i \in [n]$, and $\hat{Y} := \hat{h}_n(X)$. Note that by our definition of $\widetilde{\mathbf{A}}$, $\hat{h}_i$ is the output of $\mathbf{A}$ under the adversary sequence S_i for any $0 \leqslant i \leqslant n$. Define $\mathcal{J}_k := \{i \in [n] : K_i = k\}$ for any $k \in \mathbb{N}$. Then, for any $k \geqslant 4$, we have

$$\begin{aligned}
&\mathbb{P}(\hat{Y} \neq Y, K = k) \\
\geqslant &\mathbb{P}(\hat{Y} \neq Y, K = k,\ K_1, \dots, K_n \leqslant k,\ |\mathcal{J}_k| \leqslant \lceil k/2 \rceil,\ \hat{Y}_j \neq Y_j\ \forall j \in \mathcal{J}_k) \\
= &\mathbb{E}[\mathbb{1}\{K = k,\ K_1, \dots, K_n \leqslant k,\ |\mathcal{J}_k| \leqslant \lceil k/2 \rceil\} \\
&\quad \cdot \mathbb{P}(\hat{Y} \neq Y,\ \hat{Y}_j \neq Y_j\ \forall j \in \mathcal{J}_k \mid K, K_1, \dots, K_n, \mathcal{J}_k)] \\
\geqslant &\frac{k}{k+1}\frac{k-1}{k} \cdots \frac{k - \lceil k/2 \rceil - 1}{k - \lceil k/2 \rceil} \mathbb{P}(K = k,\ K_1, \dots, K_n \leqslant k,\ |\mathcal{J}_k| \leqslant \lceil k/2 \rceil) \\
= &\frac{k - \lceil k/2 \rceil - 1}{k+1} p_k \sum_{j=1}^{\lceil k/2 \rceil} \binom{n}{j} \left(1 - \sum_{i \geqslant k} p_i\right)^{n-j} (p_k)^j
\end{aligned}$$

$$\geqslant \frac{p_k(1-\sum_{i>k} p_i)^n}{6} \sum_{j=1}^{\lceil k/2 \rceil} \binom{n}{j} \left(\frac{1-p_k-\sum_{i>k} p_i}{1-\sum_{i>k} p_i}\right)^{n-j} \left(\frac{p_k}{1-\sum_{i>k} p_i}\right)^j.$$

Setting $n = n_i$ and $k = k_i$ for $i \geqslant 4$, by Lemma 29, we have $k_i \geqslant 4$, $n_i \geqslant 4$, $\sum_{k>k_i} p_k < 1/4n_i \leqslant 1/16$, and $4n_i p_{k_i} < k_i$. Thus, we have

$$\begin{aligned}
&\mathbb{P}(\hat{h}_{n_i}(X) \neq Y, K = k_i) \\
&\geqslant \frac{p_{k_i}(1-\sum_{k>k_i} p_k)^{n_i}}{6} \sum_{j=1}^{\lceil k_i/2 \rceil} \binom{n_i}{j} \left(\frac{1-p_{k_i}-\sum_{k>k_i} p_k}{1-\sum_{k>k_i} p_k}\right)^{n_i-j} \left(\frac{p_{k_i}}{1-\sum_{k>k_i} p_k}\right)^j \\
&\geqslant \frac{p_{k_i}(1-1/4n_i)^{n_i}}{6} \mathbb{P}(B_i \leqslant \lceil k_i/2 \rceil) \geqslant \frac{p_{k_i}(1-1/16)^4}{6} \mathbb{P}(B_i \leqslant \lceil k_i/2 \rceil),
\end{aligned}$$

where $B_i \sim \text{Binom}(n_i, q_i)$ with $q_i := p_{k_i}/(1-\sum_{k>k_i} p_k)$. Since

$$n_i q_i = \frac{n_i p_{k_i}}{1-\sum_{k>k_i} p_k} < \frac{n_i p_{k_i}}{1-1/16} < \frac{4k_i}{15},$$

by the multiplicative Chernoff bound, we have

$$\mathbb{P}(B_i \geqslant \lceil k_i/2 \rceil) \leqslant \mathbb{P}(B_i \geqslant (1+7/8)4k_i/15) \leqslant e^{-4(7/8)^2 k_i/45} \leqslant e^{-49/180}.$$

It follows that for any $i \geqslant 4$,

$$\mathbb{P}(\hat{h}_{n_i}(X) \neq Y, K = k_i) \geqslant \frac{(1-1/16)^4(1-e^{-49/180})}{6} p_{k_i} \geqslant 0.0306 C \mathcal{R}(n_i).$$

Since

$$\frac{1}{\mathcal{R}(n_i)} \mathbb{P}\left(\hat{h}_{n_i}(X) \neq Y, K = k_i \mid \boldsymbol{U}\right) \leqslant \frac{\mathbb{P}\left(K = k_i | \mid \boldsymbol{U}\right)}{\mathcal{R}(n_i)} = \frac{\mathbb{P}\left(K = k_i\right)}{\mathcal{R}(n_i)} = \frac{p_{k_i}}{R(n_i)} = C \text{ a.s.},$$

by Fatou's lemma, we have

$$\mathbb{E}\left[\limsup_{i\to\infty} \frac{\mathbb{P}(\hat{h}_{n_i}(X) \neq Y, K = k_i \mid \boldsymbol{U})}{\mathcal{R}(n_i)}\right] \geqslant \limsup_{i\to\infty} \frac{\mathbb{P}(\hat{h}_{n_i}(X) \neq Y, K = k_i)}{\mathcal{R}(n_i)} \geqslant 0.0306C.$$

Since for any $n \in \mathbb{N}$,

$$\mathbb{E}[\text{er}_{P_{\boldsymbol{U}}}(\hat{h}_n) \mid \boldsymbol{U}] = \mathbb{P}(\hat{h}_n(X) \neq Y \mid \boldsymbol{U}) \geqslant \mathbb{P}(\hat{h}_n(X) \neq Y, K = k \mid \boldsymbol{U}) \text{ a.s.},$$

we have $\mathbb{E}[\limsup_{i\to\infty} \frac{1}{\mathcal{R}(n_i)} \mathbb{E}[\text{er}_{P_{\boldsymbol{U}}}(\hat{h}_{n_i}) | \boldsymbol{U}]] \geqslant 0.0306C > 0.03C$, which implies that there exists $\boldsymbol{u} \in \mathcal{I}_\infty$ such that $\mathbb{E}[\text{er}_{P_{\boldsymbol{u}}}(\hat{h}_n)] \geqslant 0.03C\mathcal{R}(n)$ for infinitely many n. By choosing $P = P_{\boldsymbol{u}}$, we see that $\mathcal{Q}$ requires arbitrarily slow rates.

■

Lemma 29 *Let $\mathcal{R} : \mathbb{N} \to (0, 1]$ be an arbitrary rate function and $m \in \mathbb{N}$ be an arbitrary integer. Then, there exists probability masses $p_1, p_2, \cdots \geqslant 0$ so that $\sum_{i\in\mathbb{N}} p_i = 1$, two strictly increasing sequences of positive integers $(n_i)_{i\in\mathbb{N}}$ and $(k_i)_{i\in\mathbb{N}}$, and a constant $C \in [1/2\mathcal{R}(1), 1/\mathcal{R}(1))$ such that the following hold for all $i \in \mathbb{N}$:*

(a) $\sum_{k>k_i} p_k < 1/mn_i$;

(b) $mn_i p_{k_i} < k_i$;

(c) $p_{k_i} = C\mathcal{R}(n_i)$.

Proof Define $n_1 = k_1 = 1$. For any $i > 1$, define recursively

$$n_i := \inf\left\{n > n_{i-1} : \mathcal{R}(n) \leqslant \min_{j<i} \frac{\mathcal{R}(n_j)2^{j-i}}{k_j}\right\} \text{ and } k_i := \max\{m\lceil n_i\mathcal{R}(n_i)/\mathcal{R}(1)\rceil, k_{i-1}+1\}.$$

As $\mathcal{R}$ is a rate function, we have $\mathcal{R} > 0$, $\lim_{n\to\infty}\mathcal{R}(n) = 0$, and $n_i < \infty$. By construction, (n_i) and (k_i) are strictly increasing in $i \in \mathbb{N}$. Moreover, since $\mathcal{R}(n_i) \leqslant \mathcal{R}(n_1)2^{1-i}$ for any $i \in \mathbb{N}$ by definition, we have $\mathcal{R}(1) < \sum_{i\in\mathbb{N}}\mathcal{R}(n_i) \leqslant 2\mathcal{R}(1)$. Thus, we can define $C := \frac{1}{\sum_{j\in\mathbb{N}}\mathcal{R}(n_j)} \in [1/2\mathcal{R}(1), 1/\mathcal{R}(1))$. Then, for any $k \in \mathbb{N}\backslash\{k_i : i \in \mathbb{N}\}$, we define $p_k := 0$. For any $i \in \mathbb{N}$, we define $p_{k_i} := C\mathcal{R}(n_i)$. It immediately follows that $\sum_{i\in\mathbb{N}} p_i = 1$.

We now verify that the above definitions satisfy (a)-(c). For (a), since $\mathcal{R}(n_j) \leqslant \frac{\mathcal{R}(n_i)2^{i-j}}{k_i} \leqslant \frac{\mathcal{R}(1)2^{i-j}}{mn_i}$ for all $\mathbb{N} \ni i < j \in \mathbb{N}$, we have

$$\sum_{k>k_i} p_k = \sum_{j>i} C\mathcal{R}(n_j) \leqslant \frac{C\mathcal{R}(1)}{mn_i}\sum_{j>i} 2^{i-j} < \frac{1}{mn_i}.$$

For (b), we have $mn_i p_{k_i} = Cmn_i\mathcal{R}(n_i) < mn_i\mathcal{R}(n_i)/\mathcal{R}(1) \leqslant k_i$. Finally, (c) follows directly from the definition.

■

Appendix D. Exponential Upper Bound

In this section, we prove the exponential upper bound of Theorem 2 and Theorem 3. To do so, we first present a theorem on universal multiclass online learning under bandit feedback in Section D.1. Then, we prove the main result of this section in Section D.2.

D.1. Universal Multiclass Online Learning under Bandit Feedback

To prove the main theorem in this section, we utilize the result of Hanneke, Shaeiri, and Wang (2024c) on universal adversarial multiclass online learning under bandit feedback.

Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be an instance of multiclass learning under bandit feedback framework. We consider a sequential game between the learner and an adversary. At each round $t \in \mathbb{N}$, the adversary chooses an instance x_t from $\mathcal{X}$ and a label y_t from $\mathcal{Y}$ and reveals x_t to the learner. Following this, the learner predicts a label $\hat{y}_t$ from $\mathcal{Y}$. Subsequently, the learner, instead of observing the true label y_t, receives feedback $\mathbb{1}\{\hat{y}_t \neq y_t\}$, indicating whether the prediction is correct. We assume that $\mathcal{C}$ is known to the learner before starting the game. Moreover, in the realizable setting, we assume that each prefix of the sequence $\{(x_t, y_t)\}_{t=1}^{\infty}$, played by the adversary, is consistent with at least one concept in $\mathcal{C}$.

The objective is to minimize the well-known notion of the number of mistakes over time. We say that the concept class $\mathcal{C}$ is *learnable* in the multiclass online learning under bandit feedback in the realizable setting, if there is a learning algorithm that makes only finitely many mistakes in expectation on any realizable sequence played by the adversary, crucially without imposing a uniform bound on the expected number of mistakes.

The main result of Hanneke, Shaeiri, and Wang (2024c) demonstrates that the above criterion for learnability is fully characterized by the non-existence of infinite multiclass Littlestone trees. Formally, we have the following theorem.

Theorem 30 *[Hanneke, Shaeiri, and Wang (2024c)] Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. Then, we have the following dichotomy:*

- *If $\mathcal{C}$ does not have an infinite multiclass Littlestone tree, then there exists a deterministic learning algorithm that makes only finitely many mistakes against any realizable adversary.*
- *If $\mathcal{C}$ has an infinite multiclass Littlestone tree, then there is a strategy for the realizable adversary that forces any learning algorithm, including randomized, to make a linear expected number of mistakes.*

In particular, $\mathcal{C}$ is learnable in the multiclass online learning under bandit feedback in the realizable setting if and only if it has no infinite Littlestone tree.

We can verify that the algorithm of Hanneke, Shaeiri, and Wang (2024c) outputs a universally measurable function (Definition 14). In this section, we denote by $\mathbf{A}^o$ the learning algorithm based on the algorithm of Hanneke, Shaeiri, and Wang (2024c). In particular, note that both $\mathbf{A}_1^o$ and $\mathbf{A}_2^o$ are deterministic functions. Moreover, note that $\mathbf{A}_1^o$ and $\mathbf{A}_2^o$ are exactly equal.

D.2. Main Result on Learnability at an Exponential Rate

We begin with proving a critical lemma for proving the main result of the current subsection. In particular, this lemma will be useful in the proof of Theorem 32.

Lemma 31 *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework, for which $\mathcal{C}$ does not have an infinite Littlestone tree. There exist universally measurable functions $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2)$ such that $\mathbf{t}_1 : (\mathcal{X} \times \{0,1\})^\star \times \mathcal{X} \to \mathcal{Y}$ and $\mathbf{t}_2 : (\mathcal{X} \times \Sigma)^\star \to \mathbb{N}$, so that the following holds. Let $\mathcal{D}$ be a data distribution realizable by $\mathcal{C}$. Assume that there exists $t^\star \in \mathbb{N}$ such that $\mathbb{P}_{\mathcal{S}\sim\mathcal{D}^{t^\star}}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}^o}) > 0\right] \leqslant \frac{1}{10}$. For each $n \in \mathbb{N}$, when $\mathbf{t}$ is applied (in a similar fashion as Section B.2.3) to a data set $\mathcal{S} \sim \mathcal{D}^n$, it produces a data-dependent value $\hat{t}^{\mathcal{S}}_{\mathbf{t}}$, so that the following holds. There exist $C, c > 0$, possibly depending on $\mathcal{D}$ and $t^\star$, so that for every $n \in \mathbb{N}$, we have: $\mathbb{P}_{\mathcal{S}\sim\mathcal{D}^n}\{\hat{t}^{\mathcal{S}}_{\mathbf{t}} \in \mathcal{T}_{good}\} \geqslant 1 - Ce^{-cn}$, where $\mathcal{T}_{good} := \{1 \leqslant t \leqslant t^\star : \mathbb{P}_{\mathcal{S}\sim\mathcal{D}^t}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}^o}) = 0\right] \geqslant \frac{6}{10}\}$.*

Proof Fix $n \in \mathbb{N}$. Let $m_n = \lfloor (n/4)^{1/2} \rfloor$. Let $\mathcal{S} \sim \mathcal{D}^n$. First we segment the data sequence $\mathcal{S}$ into disjoint contiguous blocks, $\mathcal{A}_1, \mathcal{B}_1, \mathcal{A}_2, \mathcal{B}_2, \ldots, \mathcal{A}_{m_n}, \mathcal{B}_{m_n}$, where for every $i \in [m_n]$, $\mathcal{A}_i$ and $\mathcal{B}_i$ are of size $\lfloor n/\big(2i(i+1)\big) \rfloor$. Indeed, one can do so as we have: $n\sum_{i=1}^{m_n} \frac{1}{i(i+1)} = n\sum_{i=1}^{m_n}\left(\frac{1}{i} - \frac{1}{i+1}\right) = n\left(1 - \frac{1}{m_n+1}\right) \leqslant n$. Next, for every $i \in [m_n]$, further segment each $\mathcal{A}_i$ into disjoint contiguous batches of size i. As a result, for every $i \in [m_n]$, we have at least $M_{n,i} := \left\lfloor \frac{1}{i} \left\lfloor n/\big(2i(i+1)\big) \right\rfloor \right\rfloor \geqslant \left\lfloor n/\big(4i^2(i+1)\big) \right\rfloor$ number of batches. For every $i \in [m_n]$ and every $j \in [M_{n,i}]$, we use $\mathcal{A}_i^j$ to refer to the corresponding j^{th} batch in $\mathcal{A}_i$. Our algorithm (for defining $\hat{t}^{\mathcal{S}}_{\mathbf{t}}$) begins by running the online learning algorithm $\mathbf{A}^o$ from Theorem 30 to every batch $\mathcal{A}_i^j$, which thereby each produce a final predictor $\hat{h}^{\mathcal{A}_i^j}_{\mathbf{A}^o}$.

Subsequently, for every $i \in [m_n]$, we define:

$$b_i := \mathbb{1}\left\{ M_{n,i}^{-1} \sum_{j=1}^{M_{n,i}} \mathbb{1}\{\hat{h}^{\mathcal{A}_i^j}_{\mathbf{A}^o}(X) = Y\} < \frac{7}{10} \text{for some } (X,Y) \in \mathcal{B}_i \right\}$$

Also, for every $i \in [m_n]$, we define:

$$\hat{e}_i := \frac{1}{M_{n,i}} \sum_{j=1}^{M_{n,i}} \mathbb{1}\left\{ \hat{h}^{\mathcal{A}_i^j}_{\mathbf{A}^o}(X) \neq Y \text{ for some } (X,Y) \in \mathcal{B}_i \right\}.$$

Now, we describe the algorithm. In fact, we specify $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2)$. For each $i \in [m_n]$, we predict on all points in $\mathcal{B}_i$ using the *majority vote* prediction of $\hat{h}^{\mathcal{A}_i^1}_{\mathbf{A}^o}, \hat{h}^{\mathcal{A}_i^2}_{\mathbf{A}^o}, \ldots, \hat{h}^{\mathcal{A}_i^{M_{n,i}}}_{\mathbf{A}^o}$. Note that, based on the bandit feedback from these predictions, we are able to determine the value of b_i: that is, if the majority vote is wrong on any of the examples, $b_i = 1$ immediately, and even if the majority vote is correct on all of them, if fewer than $7/10$ of the predictors agreed with the correct prediction on any point, we still have $b_i = 1$; otherwise $b_i = 0$. Also note that, for any i with $b_i = 0$, the algorithm also has the ability to calculate $\hat{e}_i$, since the fact of $b_i = 0$ immediately implies all predictions were correct, so that we have access to all true labels in $\mathcal{B}_i$. Based on this, for any i with $b_i = 0$, we also test whether $\hat{e}_i < \frac{1}{4}$. We then define $\hat{t}$ as the minimal i that passes both of these tests: namely,

$$\hat{t} := \inf\{i \in [m_n] : b_i \neq 1 \ \wedge \ \hat{e}_i < \frac{1}{4}\}, \text{ with the convention } \inf \varnothing = \infty.$$

Now, observe that for every $i \in [m_n]$, we have:

$$\hat{e}_i \leqslant e_i := \frac{1}{M_{n,i}} \sum_{j=1}^{M_{n,i}} \mathbb{1}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{A}_i^j}_{\mathbf{A}^o}) > 0\} \quad \text{a.s.}$$

Next, consider $t^\star$ as in the statement of the lemma. By Hoeffding's inequality, there exists some $C_1, c_1 \in \mathbb{R}^+$ such that we have:

$$\begin{aligned}
\mathbb{P}[\hat{t} > t^\star] &\leqslant \mathbb{P}[b_{t^\star} = 1 \ \vee \ \hat{e}_{t^\star} \geqslant \tfrac{1}{4}] \\
&\leqslant \mathbb{P}[b_{t^\star} = 1] + \mathbb{P}[\hat{e}_{t^\star} \geqslant \tfrac{1}{4}] \\
&\leqslant \mathbb{P}[\hat{e}_{t^\star} \geqslant \tfrac{3}{10}] + \mathbb{P}[\hat{e}_{t^\star} \geqslant \tfrac{1}{4}] \\
&\leqslant \mathbb{P}[e_{t^\star} \geqslant \tfrac{3}{10}] + \mathbb{P}[e_{t^\star} \geqslant \tfrac{1}{4}] \\
&\leqslant \mathbb{P}[e_{t^\star} \geqslant \tfrac{1}{10} + \tfrac{2}{10}] + \mathbb{P}[e_{t^\star} \geqslant \tfrac{1}{10} + \tfrac{3}{20}] \\
&\leqslant \mathbb{P}[e_{t^\star} \geqslant \mathbb{E}[e_{t^\star}] + \tfrac{2}{10}] + \mathbb{P}[e_{t^\star} \geqslant \mathbb{E}[e_{t^\star}] + \tfrac{3}{20}] \\
&\leqslant \mathrm{e}^{-18\lfloor n/8(t^\star+1)^3 \rfloor/400} + \mathrm{e}^{-8\lfloor n/8(t^\star+1)^3 \rfloor/100} \\
&\leqslant C_1 \, \mathrm{e}^{-c_1 n} \, .
\end{aligned}$$

In other words, $\hat{t} \leqslant t^\star$ except with exponentially small probability.

In addition, by continuity, there exists $\varepsilon > 0$, so that for all $1 \leqslant t \leqslant t^\star$ with $t \notin \mathcal{T}_{\text{good}}$, meaning $\mathbb{P}_{\mathcal{S}' \sim \mathcal{D}^t}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}'}_{\mathbf{A}^o}) = 0\} < \frac{6}{10}$, or equivalently $\mathbb{P}_{\mathcal{S}' \sim \mathcal{D}^t}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}'}_{\mathbf{A}^o}) > 0\} > \frac{4}{10}$, we have: $\mathbb{P}_{\mathcal{S}' \sim \mathcal{D}^t}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}'}_{\mathbf{A}^o}) > \varepsilon\} > \frac{7}{20}$. Now, fix $1 \leqslant t \leqslant t^*$ with $\mathbb{P}_{\mathcal{S}' \sim \mathcal{D}^t}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}'}_{\mathbf{A}^o}) > 0\} > \frac{4}{10}$ (if such a t exists). By Hoeffding's inequality, there exist some $C_2, c_2 \in \mathbb{R}^+$ (common to all such t) such that we have:

$$\mathbb{P}\left[\frac{1}{M_{n,t}} \sum_{j=1}^{M_{n,t}} \mathbb{1}\{\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{A}_t^j}_{\mathbf{A}^o}) > \varepsilon\} < \frac{3}{10}\right] \leqslant \mathrm{e}^{-2\lfloor n/4t^2(t+1) \rfloor/200} \leqslant \mathrm{e}^{-\lfloor n/4(t+1)^3 \rfloor/100}$$

$$\leqslant \mathrm{e}^{-\lfloor n/4(t^\star+1)^3 \rfloor/100} \leqslant C_2 \, \mathrm{e}^{-c_2 n} \, .$$

Now, for every $t \in [m_n]$ if f is so that $\mathrm{er}_{\mathcal{D}}(f) > \varepsilon$ and it is independent of $\mathcal{B}_t$, we have:

$$\mathbb{P}\big\{f(X) \neq Y \text{ for some } (X,Y) \in \mathcal{B}_t\big\} \geqslant 1 - (1-\varepsilon)^{\lfloor n/(2t(t+1)) \rfloor} \leqslant 1 - (1-\varepsilon)^{\lfloor n/(2(t+1)^2) \rfloor}.$$

Thus, by applying union bound over all $j \in [M_{n,t}]$, we have that, with probability at least $1 - M_{n,t}(1-\varepsilon)^{\lfloor n/(2(t+1)^2) \rfloor}$, every $j \in [M_{n,t}]$ with $\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{A}_t^j}_{\mathbf{A}^o}) > \varepsilon$ has some $(X,Y) \in \mathcal{B}_t$ with $\hat{h}^{\mathcal{A}_t^j}_{\mathbf{A}^o}(X) \neq Y$. In particular, if this occurs together with the above event, where $> \frac{3}{10} M_{n,t}$ of these functions have error rate $> \varepsilon$, then $\hat{e}_t > \frac{3}{10}$ so that $\hat{t} \neq t$.

By the union bound over all $t \leqslant t^\star$ with $t \notin \mathcal{T}_{\text{good}}$, we have that $\hat{t}$ is not equal any of these values of t, with probability at least $1 - \sum_{t \leqslant t^\star} \left(M_{n,t}(1-\varepsilon)^{\lfloor n/(2(t+1)^2) \rfloor} + C_2 e^{-c_2 n} \right) \geqslant 1 - C_3 e^{-c_3 n}$ (for appropriate C_3, c_3).

Finally, by applying union bound combined with above results, there exist some $C, c \in \mathbb{R}^+$ such that we have:

$$\mathbb{P}\{\hat{t} \notin \mathcal{T}_{\text{good}}\} \leqslant C \, \mathrm{e}^{-cn}$$

This finishes the proof. ■

Finally, we present the main theorem of this section. The proof of this theorem is similar to the proof of Corollary 4.5 in the work of Bousquet, Hanneke, Moran, Van Handel, and Yehudayoff (2021).

Theorem 32 *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. If $\mathcal{C}$ does not have an infinite multiclass Littlestone tree, then $\mathcal{Q}$ is universal learnable under bandit feedback at rate e^{-n}.*

Proof In this proof for simplicity, we adopt the notations in the proof of Lemma 31. Before starting, note that the algorithm of Hanneke et al. (2024c) has a special property. It only changes its prediction when it makes a mistake. Thus, by combining the result from subsection D.1 with the mentioned fact, we can use the argument of Bousquet et al. (2021) in Lemma 4.3 to show the existence of a $t^\star$ satisfying the condition of Theorem 31. Now, we define the learning algorithm $\mathbf{A}$ as follows. First, we use Lemma 31, to find $\hat{t}$. This defines $\mathbf{A}_1$. Next, we use the majority vote of the functions corresponding to the batches of size $\hat{t}$ from Lemma 31 to define $\mathbf{A}_2$. Now, we prove that for every realizable distribution $\mathcal{D}$, there exist $C, c \in \mathbb{R}^+$ such that for every $n \in \mathbb{N}$, we have: $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \leqslant C\,\mathrm{e}^{-cn}$ for some constants $C, c \in \mathbb{R}^+$. To see this, note that for every $t \in \mathcal{T}_{\text{good}}$, by Hoeffding's inequality, we have: with probability less than or equal to $C_1\,\mathrm{e}^{-c_1 n}$, the majority vote of the functions corresponding to the batches of size t from Lemma 31 has non-zero error rate, for constants $C_1, c_1 \in \mathbb{R}^+$. Now, we apply union bound, thus the majority vote has zero error rate for every $t \in \mathcal{T}_{\text{good}}$ with probability at least $1 - t^\star C_1\,\mathrm{e}^{-c_1 n}$. Notice that Lemma 31 also has failure probability of less than or equal to $t^\star C_2\,\mathrm{e}^{-c_2 n}$, for constants $C_2, c_2 \in \mathbb{R}^+$. Thus, the overall failure probability is less than or equal to $C\,\mathrm{e}^{-cn}$, for constants $C, c \in \mathbb{R}^+$. Finally, note that $\mathbb{E}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}})\right] \leqslant \mathbb{P}_{\mathcal{S}\sim\mathcal{D}^n}\left[\mathrm{er}_{\mathcal{D}}(\hat{h}^{\mathcal{S}}_{\mathbf{A}}) > 0\right]$. This finishes the proof. ■

Appendix E. Linear Upper Bound

In this section, we prove the following theorem.

Theorem 33 *Let $\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ be a non-degenerate instance of multiclass learning under bandit feedback framework. If $\mathcal{C}$ has neither an infinite Natarajan Littlestone tree nor an infinite progressive Littlestone tree, then $\mathcal{Q}$ is universal learnable under bandit feedback at rate $1/n$.*

In Section E.1, we construct list learner by playing an adversarial game related to the progressive Littlestone tree. In Section E.2, we construct concept classes of bounded Natarajan dimension by playing an adversarial game related to the Natarajan Littlestone tree. In Section E.3, we introduce the one-inclusion algorithm and show its performance guarantee using concept classes constructed based on the objects studied in Section E.1 and Section E.2. We construct our final classifier and finish the proof in Section E.4.

E.1. List Learning via Progressive Littlestone Game

For any $n \in \mathbb{N}$, define $\bar{\mathcal{Y}}_n := \{(y_1, \ldots, y_n) \in \mathcal{Y}^n : |\{y_1, \ldots, y_n\}| = n\}$ to be the set of all sequences in $\mathcal{Y}^n$ with n distinct elements. Consider the following game $\mathfrak{B}$ between player $\mathtt{P}_A$ and $\mathtt{P}_L$. At each round $\tau \in \mathbb{N}$:

- Player $\mathtt{P}_A$ chooses a point $x_\tau \in \mathcal{X}$ and a sequence $\boldsymbol{y}_\tau = \{y_\tau^1, \dots, y_\tau^{\tau+1}\} \in \bar{\mathcal{Y}}_{\tau+1}$.
- Player $\mathtt{P}_L$ chooses a number $u_\tau \in [\tau+1]$.
- Player $\mathtt{P}_L$ wins the game in round τ if $\mathcal{C}_{x_1, y_1^{u_1}, \dots, x_\tau, y_\tau^{u_\tau}} = \varnothing$.

Here, for any sequence $((x_i, y_i))_{i=1}^t \in \mathcal{Z}^t$ with $t \in \mathbb{N}$,

$$\mathcal{C}_{x_1, y_1, \dots, x_\tau, y_t} := \{h \in \mathcal{C} : h(x_i) = y_i,\ \forall i \in [t]\}.$$

Define $\mathsf{P}_\infty := \prod_{t\in\mathbb{N}} \left(\mathcal{X} \times \bar{\mathcal{Y}}_{t+1} \times [t+1]\right)$ to be the set of all possible infinite sequences of plays made by $\mathtt{P}_A$ and $\mathtt{P}_L$ in $\mathfrak{B}$. We call a finite sequence of plays made by players in a game a position of the game. For any $n \in \mathbb{N}$, we define $\mathsf{P}_n := \prod_{t\in[n]} \left(\mathcal{X} \times \bar{\mathcal{Y}}_{t+1} \times [t+1]\right)$ to be the set of positions (i.e.,) of length n in $\mathfrak{B}$ and let $\mathsf{P} := \cup_{n=0}^\infty \mathsf{P}_n$ denote the set of all positions, where $\mathsf{P}_0 := \{\varnothing\}$. Since $\mathcal{X}$ is a Polish space and $\mathcal{Y}$ is countable, P_∞, P, and P_n are Polish spaces for any $n \geqslant 0$.

Let $\mathsf{W} \subseteq \mathsf{P}_\infty$ denote the set of winning sequences of $\mathtt{P}_L$. Then, we have

$$\mathsf{W} = \left\{(x_1, \boldsymbol{y}_1, u_1, \dots) \in \mathsf{P}_\infty : \exists \tau \in \mathbb{N} \text{ s.t. } \mathcal{C}_{x_1, y_1^{u_1}, \dots, x_\tau, y_\tau^{u_\tau}} = \varnothing\right\}.$$

Clearly, W is finitely decidable and thus $\mathfrak{B}$ is a Gale-Stewart game (Bousquet et al., 2021, Appendix A.1). Thus, either $\mathtt{P}_A$ or $\mathtt{P}_L$ has a winning strategy (Gale and Stewart, 1953; Kechris, 1995). Since we have assumed that $\mathcal{C}$ is a measurable concept class (Definition 12), we can show the following result.

Lemma 34 *$\mathsf{P}_\infty \backslash \mathsf{W}$ is an analytic set.*

Proof Since $\mathcal{C}$ is measurable in the sense of Definition 12, We can write

$$\begin{aligned}\mathsf{P}_\infty \backslash \mathsf{W} &= \left\{(x_1, \boldsymbol{y}_1, u_1, \dots) \in \mathsf{P}_\infty : \forall \tau \in \mathbb{N},\ \mathcal{C}_{x_1, y_1^{u_1}, \dots, x_\tau, y_\tau^{u_\tau}} \neq \varnothing\right\} \\ &= \bigcap_{\tau\in\mathbb{N}} \bigcup_{\theta\in\Theta} \bigcap_{t\in[\tau]} \{(x_1, \boldsymbol{y}_1, u_1, \dots) \in \mathsf{P}_\infty : \mathsf{h}(\theta, x_t) = y_t^{u_t}\}.\end{aligned}$$

For any $t \in \mathbb{N}$ and $i \in [t+1]$, define $U_t^i := \{(\theta, x_1, \boldsymbol{y}_1, u_1, \dots) \in \Theta \times \mathsf{P}_\infty : u_t = i\}$ which is open in the Polish space $\Theta \times \mathsf{P}_\infty$ and $A_t^i := \{(\theta, x_1, \boldsymbol{y}_1, u_1, \dots) \in \Theta \times \mathsf{P}_\infty : \mathsf{h}(\theta, x_t) = y_t^i\}$. Consider the projection $\pi_t^i : \Theta \times \mathsf{P}_\infty, (\theta, x_1, \boldsymbol{y}_1, u_1, \dots) \mapsto (\theta, x_t, y_t^i)$ which is continuous and the mapping $\widetilde{\mathsf{h}} : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}^2, (\theta, x, y) \mapsto (\mathsf{h}(\theta, x), y)$ which is Borel-measurable as h is. Since $\mathcal{Y}$ is countable, the set $\mathcal{Y}_2^= := \{(y_1, y_2) \in \mathcal{Y}^2 : y_1 = y_2\}$ is open in $\mathcal{Y}^2$. Thus, $A_t^i = (\widetilde{\mathsf{h}} \circ \pi_t^i)^{-1}(\mathcal{Y}_2^=)$ is Borel and hence $U_t^i \cap A_t^i$ is Borel. Since

$$\{(\theta, x_1, \boldsymbol{y}_1, u_1, \dots) \in \Theta \times \mathsf{P}_\infty : \mathsf{h}(\theta, x_t) = y_t^{u_t}\} = \cup_{i\in[t+1]}(U_t^i \cap A_t^i),$$

we know that $\bigcap_{t\in[\tau]}\{(\theta, x_1, \boldsymbol{y}_1, u_1, \dots) \in \Theta \times \mathsf{P}_\infty : \mathsf{h}(\theta, x_t) = y_t^{u_t}\}$ is Borel and the set

$$\bigcup_{\theta\in\Theta} \bigcap_{t\in[\tau]} \{(x_1, \boldsymbol{y}_1, u_1, \dots) \in \mathsf{P}_\infty : \mathsf{h}(\theta, x_t) = y_t^{u_t}\},$$

as a projection of a Borel set, is analytic. Finally, we can conclude that $\mathsf{P}_\infty \backslash \mathsf{W}$ is is an analytic set. ■

We next prove that the existence of a winning strategy for $\mathtt{P}_A$ is equivalent to the existence of an infinite progressive Littlestone tree of $\mathcal{C}$.

Lemma 35 *$\mathtt{P}_A$ has a winning strategy in $\mathfrak{B}$ if and only if $\mathcal{C}$ has an infinite progressive Littlestone tree.*

Proof For any $k \in \{0\} \cup \mathbb{N}$, define $\mathcal{I}_k := [2] \times [3] \times \cdots \times [k+1]$ with the convention that $\mathcal{I}_0 = \varnothing$. Define $\mathcal{I} := \cup_{k=0}^{\infty} \mathcal{I}_k$.

If $\mathtt{P}_A$ has a winning strategy in $\mathfrak{B}$, denote the strategy by $\xi : \mathcal{I} \to \cup_{t \in \mathbb{N}} (\mathcal{X} \times \bar{\mathcal{Y}}_{t+1})$ which for each $t \in \{0\} \cup \mathbb{N}$, given the choices $(u_1, \ldots, u_t) \in \mathcal{I}_t$ of $\mathtt{P}_L$ up to round t, outputs $\xi(u_1, \ldots, u_t) = (x, \boldsymbol{y}) \in \mathcal{X} \times \bar{\mathcal{Y}}_{t+2}$. For convenience, we use $\xi(u_1, \ldots, u_t)(1)$ to denote the output instance $x \in \mathcal{X}$ and use $\xi(u_1, \ldots, u_t)(2)$ to denote the output sequence $\boldsymbol{y} \in \bar{\mathcal{Y}}_{t+2}$. Then, ξ naturally defines an infinite progressive $(\mathcal{X}, \mathcal{Y})$-valued tree $\mathcal{T}$ (Definition 20) represented as

$$\mathcal{T} = \{x_{\boldsymbol{u}} = \xi(\boldsymbol{u})(1) : \boldsymbol{u} \in \mathcal{I}\} \cup \{\boldsymbol{y}_{\boldsymbol{u}} = \xi(\boldsymbol{u})(2) : \boldsymbol{u} \in \mathcal{I}\},$$

where for $x_{\varnothing}$ (when $\boldsymbol{u} = \varnothing$) is the label of the root node $v_{\varnothing}$ and $\boldsymbol{y}_{\varnothing} \in \bar{\mathcal{Y}}_2$ is the sequence of labels of the edges between $v_{\varnothing}$ and its two children. Then, inductively, given any $\boldsymbol{u} = (u_1, \ldots, u_{t-1}) \in \mathcal{I}_{t-1}$ and $u_t \in [t+1]$ with $t \in \mathbb{N}$, we let $v_{(\boldsymbol{u}, u_t)}$ denote the u_t-th child of $v_{\boldsymbol{u}}$, $x_{(\boldsymbol{u}, u_t)}$ be the label of $v_{(\boldsymbol{u}, u_t)}$, and $\boldsymbol{y}_{(\boldsymbol{u}, u_t)}$ be the sequence of the labels between $v_{(\boldsymbol{u}, u_t)}$ and its children. Since ξ is a winning strategy of $\mathtt{P}_A$, $\mathcal{T}$ is shattered by $\mathcal{C}$. Thus, $\mathcal{C}$ has an infinite progressive Littlestone tree $\mathcal{T}$.

On the other hand, if $\mathcal{C}$ has an infinite progressive Littlestone tree

$$\mathcal{T} = \{x_{\boldsymbol{u}} : \boldsymbol{u} \in \mathcal{I}\} \cup \{\boldsymbol{y}_{\boldsymbol{u}} : \boldsymbol{u} \in \mathcal{I}\},$$

where for each $\boldsymbol{u} \in \mathcal{I}$, $x_{\boldsymbol{u}} \in \mathcal{X}$ and $\boldsymbol{y}_{\boldsymbol{u}} \in \bar{\mathcal{Y}}_{|\boldsymbol{u}|+2}$, then $\mathcal{T}$ naturally defines a strategy $\xi : \mathcal{I} \to \cup_{t \in \mathbb{N}} (\mathcal{X} \times \bar{\mathcal{Y}}_{t+1})$, $\boldsymbol{u} \mapsto (x_{\boldsymbol{u}}, \boldsymbol{y}_{\boldsymbol{u}}) \in \mathcal{X} \times \bar{\mathcal{Y}}_{|\boldsymbol{u}|+2}$ for $\mathtt{P}_A$. Since $\mathcal{T}$ is shattered by $\mathcal{C}$, ξ is winning for $\mathtt{P}_A$. Thus, $\mathtt{P}_A$ has a winning strategy. ■

Furthermore, Bousquet et al. (2021, Appendix B) constructs a universally measurable winning strategy for $\mathtt{P}_L$ in a Gale-Stewart game provided that $\mathsf{P}_\infty \backslash \mathsf{W}$ is an analytic set. Thus, we have the following corollary from Lemma 34, Lemma 35, and Bousquet et al. (2021, Theorem B.1).

Corollary 36 *If $\mathcal{C}$ does not have an infinite progressive Littlestone tree, then $\mathtt{P}_L$ has a universally measurable winning strategy in $\mathfrak{B}$.*

Now we assume that $\mathcal{C}$ does not have an infinite progressive Littlestone tree. By Bousquet et al. (2021, Definition B.5), we can define the game value $\mathrm{val} : \mathsf{P} \to \mathrm{ORD}^*$ for the Gale-Stewart game $\mathfrak{B}$. Define the following online algorithm which runs on a sequence of feature-label pairs $((x_1, y_1), (x_2, y_2), \ldots) \in \mathcal{Z}^{\mathbb{N}}$.

1. Initialize $\tau_0 \leftarrow 1$ and $\boldsymbol{v}_0 \leftarrow \varnothing$.
2. For each round $t \in \mathbb{N}$:
   (a) Let $\tau_t \leftarrow \tau_{t-1}$ and $\boldsymbol{v}_t \leftarrow \boldsymbol{v}_{t-1}$.
   (b) For each $\boldsymbol{y} = (y^1, \ldots, y^{\tau_{t-1}+1}) \in \bar{\mathcal{Y}}_{\tau_{t-1}+1}$ such that $y_t \in \boldsymbol{y}$:

i. Let $u \in [\tau_{t-1}+1]$ be such that $y_t = y^u$.
ii. If $\mathrm{val}(\boldsymbol{v}_{t-1}, x_t, \boldsymbol{y}, u) < \min\{\mathrm{val}(\boldsymbol{v}_{t-1}), \mathrm{val}(\varnothing)\}$:
   A. Let $\bar{x}_{\tau_{t-1}} \leftarrow x_t$, $\boldsymbol{y}_{\tau_{t-1}} \leftarrow \boldsymbol{y}$, $u_{\tau_{t-1}} \leftarrow u$, $\boldsymbol{v}_t \leftarrow (\boldsymbol{v}_{t-1}, \bar{x}_{\tau_{t-1}}, \boldsymbol{y}_{\tau_{t-1}}, u_{\tau_{t-1}})$ and $\tau_t \leftarrow \tau_{t-1}+1$.
   B. Break.

By the above algorithm, we have define the following mappings for each $t \in \mathbb{N}$,

$$T_{t-1} : \mathcal{Z}^{t-1} \to [t],\ ((x_1,y_1),\dots,(x_{t-1},y_{t-1})) \mapsto \tau_{t-1} \text{ and}$$
$$\boldsymbol{V}_{t-1} : \mathcal{Z}^{t-1} \to \cup_{i=0}^{t-1}\mathsf{P}_i,\ ((x_1,y_1),\dots,(x_{t-1},y_{t-1})) \mapsto \boldsymbol{v}_{t-1}$$

with the convention that $\mathcal{Z}^0 = \{\varnothing\}$. For notational convenience, we define the function $\mathfrak{u} : \cup_{t\in\mathbb{N}}(\bar{\mathcal{Y}}_{t+1} \times [t+1]) \to \cup_{t\in\mathbb{N}}(\bar{\mathcal{Y}}_{t+1} \times \mathcal{Y})$ which for any $t \in \mathbb{N}$, $\boldsymbol{y} = (y^1,\dots,y^{t+1}) \in \bar{\mathcal{Y}}_{t+1}$, and $u \in [t+1]$, maps $(\boldsymbol{y}, u)$ to $(\boldsymbol{y}, y^u)$. It is obvious that $\mathfrak{u}$ is an injective function. Now, for any $t \in \mathbb{N} \cup \{0\}$, we can define the mapping

$$\boldsymbol{L}_t : \mathcal{Z}^t \times \mathcal{X} \to 2^{\mathcal{Y}},\ (\boldsymbol{z}, x) \mapsto \big\{y \in \mathcal{Y} : \mathrm{val}(\boldsymbol{V}_t(\boldsymbol{z}), x, \mathfrak{u}^{-1}(\boldsymbol{y}, y)) \geqslant \mathrm{val}(\boldsymbol{V}_t(\boldsymbol{z})), \\ \forall \boldsymbol{y} \in \bar{\mathcal{Y}}_{T_t(\boldsymbol{z})+1} \text{ with } \boldsymbol{y} \ni y\big\},$$

where $\boldsymbol{z} \in \mathcal{Z}^t$ and $x \in \mathcal{X}$. By (Bousquet et al., 2021, Remark 5.4 and Appendix B.4), T_{t-1}, $\boldsymbol{V}_{t-1}$, and $\boldsymbol{L}_{t-1}$ are universally measurable for all $t \in \mathbb{N}$. We show the following results about $\boldsymbol{L}_t$.

Lemma 37 *Let $\boldsymbol{z} \in \mathcal{Z}^t$ be a $\mathcal{C}$-realizable sequence with $t \in \mathbb{N}$. Then, for any $x \in \mathcal{X}$, we have $|\boldsymbol{L}_t(\boldsymbol{z}, x)| \leqslant T_t(\boldsymbol{z})$.*

Proof Define $\boldsymbol{v} := \boldsymbol{V}_t(\boldsymbol{z})$ and $\tau := T_t(\boldsymbol{z})$. Since $\mathcal{C}$ does not have an infinite progressive Littlestone tree, by Lemma 35, we have $\mathrm{val}(\boldsymbol{w}) < \Omega$ for any $\boldsymbol{w} \in \mathsf{P}$. Since $\boldsymbol{z}$ is $\mathcal{C}$-realizable, the definition of $\boldsymbol{V}_t$ ensures that $\mathrm{val}(\boldsymbol{v}) \geqslant 0$. Suppose on the contrary that $|\boldsymbol{L}_t(\boldsymbol{z}, x)| \geqslant \tau+1$, pick arbitrary $\tau+1$ of the elements in $\boldsymbol{L}_t(\boldsymbol{z}, x)$ to form a sequence $\boldsymbol{y} \in \bar{\mathcal{Y}}_{\tau+1}$. Then, for any $u \in [\tau+1]$, we have $\mathrm{val}(\boldsymbol{v}, x, \boldsymbol{y}, u) \geqslant \mathrm{val}(\boldsymbol{v})$, which contradicts Bousquet et al. (2021, Proposition B.8) as we have shown that $0 \leqslant \mathrm{val}(\boldsymbol{v}) < \Omega$. Thus, it must be the case that $|\boldsymbol{L}_t(\boldsymbol{z}, x)| \leqslant \tau$. ■

Recall that an infinite sequence $((x_1,y_1),(x_2,y_2),\dots) \in \mathcal{Z}^{\mathbb{N}}$ is $\mathcal{C}$-realizable if for any $t \in \mathbb{N}$, there exists some $h \in \mathcal{C}$ such that $h(x_i) = y_i$ for all $i \in [t]$.

Lemma 38 *Let $\boldsymbol{z} = ((x_1,y_1),(x_2,y_2),\dots) \in \mathcal{Z}^{\mathbb{N}}$ be a $\mathcal{C}$-realizable sequence and let $\boldsymbol{z}_t$ denote $((x_1,y_1),\dots,(x_t,y_t))$ for each $t \in \mathbb{N}$. Then, there exists some $t_{\boldsymbol{z}} \in \mathbb{N}$ such that $y_t \in \boldsymbol{L}_{t-1}(\boldsymbol{z}_{t-1}, x_t)$ for all $t \geqslant t_{\boldsymbol{z}}$. It immediately follows that $T_t(\boldsymbol{z}_t) = T_{t-1}(\boldsymbol{z}_{t-1})$, $\boldsymbol{V}_t(\boldsymbol{z}_t) = \boldsymbol{V}_{t-1}(\boldsymbol{z}_{t-1})$, and $\boldsymbol{L}_t(\boldsymbol{z}_t, \cdot) = \boldsymbol{L}_{t-1}(\boldsymbol{z}_{t-1}, \cdot)$ for all $t \geqslant t_{\boldsymbol{z}}$.*

Proof Let $\tau_{t-1} = T_{t-1}(\boldsymbol{z}_{t-1})$ and $\boldsymbol{v}_{t-1} = \boldsymbol{V}_{t-1}(\boldsymbol{z}_{t-1})$ for each $t \in \mathbb{N}$. Define $t_0 := 0$ and $t_1 := \inf\{t \geqslant 1 : \tau_t = \tau_{t-1}+1\}$ with the convention that $\inf\varnothing = \infty$. Suppose that we have defined $t_0, t_1, \dots, t_i$ for some $i \in \mathbb{N}$. Then, we define

$$t_{i+1} := \begin{cases} \inf\{t \geqslant t_i+1 : \tau_t = \tau_{t-1}+1\} & \text{if } t_i < \infty, \\ \infty & \text{if } t_i = \infty. \end{cases}$$

By induction, we have defined $t_i \in \mathbb{N} \cup \{\infty\}$ for all $i \in \mathbb{N}$. Define $n := \inf\{i \in \mathbb{N} : t_i = \infty\} \in \mathbb{N} \cup \{\infty\}$, $\boldsymbol{w}_0 := \boldsymbol{v}_0 = \varnothing$, and $\boldsymbol{w}_i := \boldsymbol{v}_{t_i}$ for all $1 \leqslant i < n$. Then for each $1 \leqslant i < n$, the inequality in step 2(b)ii of the online algorithm is true in round t_i and is false in round t for $t_{i-1} < t < t_i$, which implies that (following the notation in the algorithm) for $\bar{x}_i := x_{t_i} \in \mathcal{X}$, there exists $\boldsymbol{y}_i = (y_i^1, \dots, y_i^{i+1}) \in \bar{\mathcal{Y}}_{i+1}$ and $u_i \in [i+1]$ such that $y_{t_i} = y_i^{u_i}$, $\boldsymbol{w}_i = (\boldsymbol{w}_{i-1}, \bar{x}_i, \boldsymbol{y}_i, u_i)$, and

$$\operatorname{val}(\boldsymbol{w}_i) < \min\{\operatorname{val}(\boldsymbol{w}_{i-1}), \operatorname{val}(\varnothing)\}. \tag{1}$$

We next show that $\operatorname{val}(\boldsymbol{w}_i) < \operatorname{val}(\boldsymbol{w}_{i-1}) \leqslant \operatorname{val}(\varnothing)$ for all $1 \leqslant i < n$ by induction. Since $\boldsymbol{w}_0 = \varnothing$, we have $\operatorname{val}(\boldsymbol{w}_1) < \operatorname{val}(\boldsymbol{w}_0) = \operatorname{val}(\varnothing)$ by (1). Suppose that for some $i \in \mathbb{N}$ with $i + 1 < n$, we have $\operatorname{val}(\boldsymbol{w}_j) < \operatorname{val}(\boldsymbol{w}_{j-1}) \leqslant \operatorname{val}(\varnothing)$ for all $j \in [i]$. Then, (1) implies that $\operatorname{val}(\boldsymbol{w}_{i+1}) < \operatorname{val}(\boldsymbol{w}_i)$. Thus, we have $\operatorname{val}(\boldsymbol{w}_i) < \operatorname{val}(\boldsymbol{w}_{i-1}) \leqslant \operatorname{val}(\varnothing)$ for all $1 \leqslant i < n$. Moreover, since $\mathcal{Q}$ does not have an infinite progressive Littlestone tree, by Lemma 35, we have $\operatorname{val}(\varnothing) < \Omega$. Actually, since $\mathsf{P}_\infty \backslash \mathsf{W}$ is analytic by Lemma 34, we further have $\operatorname{val}(\varnothing) < \omega_1$ (Bousquet et al., 2021, Lemma B.7).

On the other hand, since $\boldsymbol{z}$ is $\mathcal{C}$-realizable, for any $i \in \mathbb{N}$ with $i < n$, there exists $h_i \in \mathcal{C}$ such that $h_i(x_{t_j}) = y_{t_j}$ for all $j \in [i]$. It follows that $h_i(\bar{x}_j) = y_j^{u_j}$ for all $j \in [i]$, $\mathcal{C}_{\bar{x}_1, y_1^{u_1}, \dots, \bar{x}_i, y_i^{u_i}} \neq \varnothing$, and $\operatorname{val}(\boldsymbol{w}_i) \geqslant 0$. In summary, we have proved that for all $i \in \mathbb{N}$ with $i < n$, we have

$$0 \leqslant \operatorname{val}(\boldsymbol{w}_i) < \operatorname{val}(\boldsymbol{w}_{i-1}) \leqslant \operatorname{val}(\varnothing) < \omega_1. \tag{2}$$

Suppose for contradiction that $n = \infty$. Then by (2), we have obtained an infinite strictly decreasing sequence of ordinals, contradicting the well-ordering of ordinals. Thus, we must have $n < \infty$. Now, setting $t_{\boldsymbol{z}} = t_{n-1} + 1 \in \mathbb{N}$ completes the proof. ∎

The following corollary follows directly from Lemma 37 and Lemma 38.

Corollary 39 *Let $\boldsymbol{z} \in \mathcal{Z}^{\mathbb{N}}$ be a $\mathcal{C}$-realizable sequence and let $\boldsymbol{z}_t$ denote $\boldsymbol{z}_{\leqslant t}$ for each $t \in \mathbb{N}$. Then, there exists some $t_{\boldsymbol{z}} \in \mathbb{N}$ and $\tau_{\boldsymbol{z}} := T_{t_{\boldsymbol{z}}-1}(\boldsymbol{z}_{t_{\boldsymbol{z}}-1}) \leqslant t_{\boldsymbol{z}}$ such that $y \in \boldsymbol{L}_{t-1}(\boldsymbol{z}_{t-1}, x)$ and $|\boldsymbol{L}_{t-1}(\boldsymbol{z}_{t-1}, x)| \leqslant \tau_{\boldsymbol{z}}$ for any $t \geqslant t_{\boldsymbol{z}}$ and any $(x, y) \in \mathcal{Z}$ satisfying that $(\boldsymbol{z}_{t-1}, x, y)$ is $\mathcal{C}$-realizable.*

For any menu $\mu : \mathcal{X} \to 2^{\mathcal{Y}}$ and distribution P over $\mathcal{X} \times \mathcal{Y}$, we use $\operatorname{er}_P(\mu) := P\{(x, y) \in \mathcal{Z} : y \notin \mu(x)\}$ to denote the error rate of μ under P. Let P be a $\mathcal{C}$-realizable distribution. For any $\boldsymbol{z} \in \mathcal{Z}^t$ with $t \in \mathbb{N} \cup \{0\}$, define the menu $\mu_{\boldsymbol{z}} : X \to 2^{\mathcal{Y}}$, $x \mapsto \boldsymbol{L}_t(\boldsymbol{z}, x)$. For any menu $\mu : \mathcal{X} \to 2^{\mathcal{Y}}$, we use $|\mu| := \sup_{x \in \mathcal{X}} |\mu(x)|$ to denote its size. Then, we have the following guarantee.

Lemma 40 *Sample $S = ((X_1, Y_1), (X_2, Y_2), \dots) \sim P^{\mathbb{N}}$. For any $t \in \{0\} \cup \mathbb{N}$, let S_t denote $S_{\leqslant t}$. Then, there exists a finite random variable $N \in \mathbb{N}$ such that for all $t \in \mathbb{N}$, if $\operatorname{er}_P(\mu_{S_j}) > 0$ or $|\mu_{S_j}| > t$ for some $j \geqslant t$, then $N > t$, almost surely. It follows that $\lim_{t \to \infty} \mathbb{P}(\exists j \geqslant t \text{ s.t. } \operatorname{er}_P(\mu_{S_j}) > 0 \text{ or } |\mu_{S_j}| > t) = 0$.*

Proof Since $P \in \mathrm{RE}(\mathcal{C})$, S is $\mathcal{C}$-realizable almost surely. For any $t \in \mathbb{N}$, since S_t and $((X_{t+1}, Y_{t+1}), (X_{t+2}, Y_{t+2}), \dots)$ are independent, by the strong laws of large number, we have

$$\lim_{m \to \infty} \frac{1}{m} \sum_{i=t+1}^{t+m} \mathbb{1}\{Y_i \notin \mu_{S_t}(X_i)\} = \operatorname{er}_P(\mu_{S_t}). \tag{3}$$

By Lemma 37 and Lemma 38, with probability one, there exists a finite $N \in \mathbb{N}$ such that for any $i \geqslant N$, we have $\mu_{S_i} = \mu_{S_{N-1}}$, $|\mu_{S_{N-1}}| \leqslant T_{N-1}(S_{N-1}) \leqslant N$, and $Y_i \in \mu_{S_{i-1}}(X_i) = \mu_{S_{N-1}}(X_i)$. Thus, $t \geqslant N$ implies that for all $j \geqslant t$, $|\mu_{S_j}| \leqslant N \leqslant t$ and $Y_i \in \mu_{S_{N-1}}(X_i) = \mu_{S_j}(X_i)$ for all $i \geqslant t$, which further implies that $\mathrm{er}_P(\mu_{S_j}) = 0$ by (3).

In conclusion, for all $t \in \mathbb{N}$, the existence of some $j \geqslant t$ such that $\mathrm{er}_P(\mu_{S_j}) > 0$ or $|\mu_{S_j}| > t$ implies that $N > t$, almost surely. It follows that

$$\mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_P(\mu_{S_j}) > 0 \text{ or } |\mu_{S_j}| > t) \leqslant \mathbb{P}(N > t).$$

Since $N < \infty$, we have $\lim_{t\to\infty} \mathbb{P}(N > t) = 0$ and

$$\lim_{t\to\infty} \mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_P(\mu_{S_j}) > 0 \text{ or } |\mu_{S_j}| > t) = 0.$$

■

Since we work with bandit feedback, we first prove the following result which shows that we are able to "guess" correctly infinitely many labels.

Lemma 41 *Let D_1 and D_2 be two probability measures on $\mathcal{Y}$ such that*

$$p := \textstyle\sum_{y\in\mathcal{Y}} D_1(\{y\})D_2(\{y\}) > 0.$$

Sample $(Y_1^1, Y_2^1, \dots) \sim (D_1)^{\mathbb{N}}$ and $(Y_1^2, Y_2^2, \dots) \sim (D_2)^{\mathbb{N}}$ independently. Then, there exists an infinite increasing sequence of positive integers $1 \leqslant I_1 < I_2 < \cdots$ such that $Y_{I_j}^1 = Y_{I_j}^2$ for all $j \in \mathbb{N}$ almost surely.

Proof For any $i \in \mathbb{N}$, the events $\{Y_1^1 = Y_1^2\}, \{Y_2^1 = Y_2^2\}, \{Y_3^1 = Y_3^2\}, \dots$ are independent and

$$\sum_{i=1}^{\infty} \mathbb{P}(Y_i^1 = Y_i^2) = \sum_{i=1}^{\infty} \sum_{y\in\mathcal{Y}} \mathbb{P}(Y_i^1 = y, Y_i^2 = y) = \sum_{i=1}^{\infty} \sum_{y\in\mathcal{Y}} D_1(\{y\})D_2(\{y\}) = \sum_{i=1}^{\infty} p = \infty.$$

Then, by the second Borel-Cantelli lemma, we know that $Y_i^1 = Y_i^2$ for infinitely many i's almost surely. ■

Let D be a probability measure on $\mathcal{Y}$ such that $D(y) := D(\{y\}) > 0$ for all $y \in \mathcal{Y}$. Define the mapping $\hat{\boldsymbol{z}} : \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t) \to \mathcal{Z}^*$ such that $\hat{\boldsymbol{z}}(\varnothing) := \varnothing$ and for any $t \in \mathbb{N}$, $\boldsymbol{z} = ((x_1, y_1), \dots, (x_t, y_t)) \in \mathcal{Z}^t$, and $\boldsymbol{y} = (y_1', \dots, y_t') \in \mathcal{Y}^t$, let $\hat{\boldsymbol{z}}(\boldsymbol{z}, \boldsymbol{y}) := ((x_i, y_i) : i \in [t], y_i = y_i')$. We can show the following result about $\hat{\boldsymbol{z}}$.

Lemma 42 *Sample $S = ((X_1, Y_1), (X_2, Y_2), \dots) \sim P^{\mathbb{N}}$ and $\boldsymbol{Y} = (Y_1', Y_2', \dots) \sim D^{\mathbb{N}}$ independently. Then, we have $\boldsymbol{Z} := \hat{\boldsymbol{z}}(S, \boldsymbol{Y}) \sim (P_D)^{\mathbb{N}}$, where P_D a distribution over $\mathcal{Z}$ defined by $P_D(E) := \mathbb{P}((X_1, Y_1) \in E \mid Y_1 = Y_1')$ for any measurable subset E of $\mathcal{Z}$. Moreover, if $P \in \mathrm{RE}(\mathcal{C})$, then $P_D \in \mathrm{RE}(\mathcal{C})$. For any $\mu : \mathcal{X} \to 2^{\mathcal{Y}}$, if $\mathrm{er}_P(\mu) > 0$, then $\mathrm{er}_{P_D}(\mu) > 0$.*

Proof Define $p_{P,D} := \mathbb{P}(Y = Y') = \sum_{y\in\mathcal{Y}} P(\mathcal{X}\times\{y\})D(y) > 0$. For any $h : \mathcal{X} \to \mathcal{Y}$, we have

$$\mathrm{er}_{P_D}(h) = \frac{\mathbb{P}(h(X_1) \neq Y_1, Y_1 = Y_1')}{\mathbb{P}(Y_1 = Y_1')} \leqslant \frac{\mathbb{P}(h(X_1) \neq Y_1)}{\mathbb{P}(Y_1 = Y_1')} = \frac{\mathrm{er}_P(h)}{p_{P,D}}.$$

Thus, if $P \in \mathrm{RE}(\mathcal{C})$, then we have $\inf_{h\in\mathcal{C}} \mathrm{er}_{P_D}(h) = 0$ and $P_D \in \mathrm{RE}(\mathcal{C})$. By Lemma 41, $\boldsymbol{Z}$ is an infinite sequence. For any $n \in \mathbb{N}$ and measurable sets $E_1, \dots, E_n \subseteq \mathcal{Z}$, we have

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Z}_{\leqslant n} \in E_1 \times \cdots \times E_n)\\
=& \sum_{\boldsymbol{i}=(i_1,\dots,i_n):1\leqslant i_1<i_2<\cdots<i_n} \mathbb{P}((X_{i_j}, Y_{i_j}) \in E_j, Y_{i_j} = Y_{i_j}',\ \forall j \in [n],\ Y_t \neq Y_t',\ \forall t \in [i_n]\backslash\boldsymbol{i})\\
=& \sum_{\boldsymbol{i}=(i_1,\dots,i_n):1\leqslant i_1<i_2<\cdots<i_n} \Big(\prod_{j\in[n]} \mathbb{P}((X_{i_j}, Y_{i_j}) \in E_j, Y_{i_j} = Y_{i_j}')\Big)\Big(\prod_{t\in[i_n]\backslash\boldsymbol{i}} \mathbb{P}(Y_t \neq Y_t')\Big)\\
=& \sum_{\boldsymbol{i}=(i_1,\dots,i_n):1\leqslant i_1<i_2<\cdots<i_n} p_{P,D}^n (1-p_{P,D})^{i_n-n} \prod_{j\in[n]} P_D(E_j)\\
=& \prod_{j\in[n]} P_D(E_j) \sum_{\boldsymbol{i}=(i_1,\dots,i_n):1\leqslant i_1<i_2<\cdots<i_n} \mathbb{P}(Y_{i_j} = Y_{i_j}',\ \forall j \in [n],\ Y_t \neq Y_t',\ \forall t \in [i_n]\backslash\boldsymbol{i})\\
=& \prod_{j\in[n]} P_D(E_j)\mathbb{P}(|\boldsymbol{Z}| \geqslant n)\\
=& \prod_{j\in[n]} P_D(E_j),
\end{aligned}$$

where the last equality follows from the fact that $|\boldsymbol{Z}|$ is infinite. Thus, $\boldsymbol{Z} \sim (P_D)^{\mathbb{N}}$.

If $\mathrm{er}_P(\mu) > 0$, since $\mathrm{er}_P(\mu) = \sum_{y'\in\mathcal{Y}} P(\{(x,y) \in \mathcal{Z} : y' = y \notin \mu(x)\})$, there exists some $y' \in \mathcal{Y}$ such that $P(\{(x,y) \in \mathcal{Z} : y' = y \notin \mu(x)\}) > 0$. It follows that

$$\begin{aligned}
\mathrm{er}_{P_D}(\mu) =& \frac{\sum_{y''\in\mathcal{Y}} P(\{(x,y) \in \mathcal{Z} : y'' = y \notin \mu(x)\})D(y'')}{p_{P,D}}\\
\geqslant& \frac{P(\{(x,y) \in \mathcal{Z} : y' = y \notin \mu(x)\})D(y')}{p_{P,D}} > 0.
\end{aligned}$$

■

Now, we can define

$$\hat{\boldsymbol{L}} : \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t \times \mathcal{X}) \to 2^{\mathcal{Y}},\ (\boldsymbol{z}, \boldsymbol{y}, x) \mapsto \boldsymbol{L}_{|\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})|}(\hat{\boldsymbol{z}}(\boldsymbol{z}, \boldsymbol{y}), x).$$

For notational convenience, we also define $\hat{\mu}_{\boldsymbol{z},\boldsymbol{y}} : \mathcal{X} \to 2^{\mathcal{Y}}$, $x \mapsto \hat{\boldsymbol{L}}(\boldsymbol{z}, \boldsymbol{y}, x)$ for any $(\boldsymbol{z}, \boldsymbol{y}) \in \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t)$. Since $\hat{\boldsymbol{z}}$ is measurable and $\boldsymbol{L}$ is universally measurable, $\hat{\boldsymbol{L}}$ is also universally measurable. We can show the following result for $\hat{\boldsymbol{L}}$ based on Lemma 40.

Lemma 43 *Sample $S = ((X_1, Y_1), (X_2, Y_2), \dots) \sim P^{\mathbb{N}}$ and $\boldsymbol{Y} = (Y_1', Y_2', \dots) \sim D^{\mathbb{N}}$ independently. For any $t \in \{0\} \cup \mathbb{N}$, let S_t denote $S_{\leqslant t}$ and $\boldsymbol{Y}_t$ denote $\boldsymbol{Y}_{\leqslant t}$. Then, we have*

$$\lim_{t\to\infty} \mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_P(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t) = 0.$$

Proof Let P_D denote the distribution over $\mathcal{Z}$ specified in Lemma 42. Then, we have $P_D(E) = \mathbb{P}((X_1, Y_1) \in E \mid Y_1 = Y_1')$ for any measurable subset E of $\mathcal{Z}$. By Lemma 42, since $P \in \mathrm{RE}(\mathcal{C})$, we have $P_D \in \mathrm{RE}(\mathcal{C})$ and $\boldsymbol{Z} := \hat{\boldsymbol{z}}(S, \boldsymbol{Y}) \sim (P_D)^{\mathbb{N}}$. Define $K_t := |\hat{\boldsymbol{z}}(S_t, \boldsymbol{Y}_t)|$ for any $t \in \mathbb{N} \cup \{0\}$. Then, we have $\lim_{t\to\infty} K_t = \infty$, almost surely.

Applying Lemma 40 for $\boldsymbol{Z} \sim (P_D)^{\mathbb{N}}$, we know that there exists a finite random random variable $N \in \mathbb{N}$ such that for all $t \in \mathbb{N}$, the existence of some $j \geqslant t$ for which $\mathrm{er}_{P_D}(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0$ or $|\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > K_t$ implies that $N > K_t$, almost surely. Since $K_t \leqslant t$, we have

$$\begin{aligned}&\mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_{P_D}(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t)\\ \leqslant&\mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_{P_D}(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > K_t) \leqslant \mathbb{P}(N > K_t),\end{aligned}$$

Since $\lim_{t\to\infty} K_t = \infty$, we have $K_t > N$ for all sufficiently large t almost surely. It follows that $\lim_{t\to\infty} \mathbb{P}(N > K_t) = 0$ and

$$\lim_{t\to\infty} \mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_{P_D}(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t) = 0.$$

Since by Lemma 42, $\mathrm{er}_P(\mu) > 0$ implies that $\mathrm{er}_{P_D}(\mu) > 0$ for any $\mu : \mathcal{X} \to 2^{\mathcal{Y}}$, we have

$$\mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_P(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t) \leqslant \mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_{P_D}(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t)$$

and

$$\lim_{t\to\infty} \mathbb{P}(\exists j \geqslant t \text{ s.t. } \mathrm{er}_P(\hat{\mu}_{S_j,\boldsymbol{Y}_j}) > 0 \text{ or } |\hat{\mu}_{S_j,\boldsymbol{Y}_j}| > t) = 0.$$

■

E.2. Concept Classes of Bounded Natarajan Dimension via Natarajan Littlestone Game

Now assume that $\mathcal{C}$ does not have an infinite Natarajan Littlestone tree. For completeness, we include the results from Kalavasis et al. (2022) on constructing concept classes of bounded Natarajan dimension, by proceeding the Natarajan Littlestone game $\mathscr{B}$ between players $\mathtt{P}_A$ and $\mathtt{P}_L$ defined below. Before presenting the game, we first define a mapping $\boldsymbol{r} : \cup_{t=0}^{\infty}(\mathcal{Y}^{2t} \times \{0,1\}^t) \to \cup_{t=0}^{\infty}\mathcal{Y}^t$ which for $t = 0$, maps $\varnothing$ to $\varnothing$ and for $t \in \mathbb{N}$, maps $\boldsymbol{y}^{(0)} = (y^{(0)1}, \ldots, y^{(0)t}) \in \mathcal{Y}^t$, $\boldsymbol{y}^{(1)} = (y^{(1)1}, \ldots, y^{(1)t}) \in \mathcal{Y}^t$, and $\boldsymbol{u} = (u^1, \ldots, u^t) \in \{0,1\}^t$ to

$$\boldsymbol{r}(\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \boldsymbol{u}) := (y^1, \ldots, y^t) \text{ with } y^i = y^{(u^i)i},\ \forall i \in [t].$$

For the game $\mathscr{B}$, at each round $\tau \in \mathbb{N}$:

- Player $\mathtt{P}_A$ chooses a sequence of points $\boldsymbol{x}_\tau = (x_\tau^1, \ldots, x_\tau^\tau) \in \mathcal{X}^\tau$ and two sequences of everywhere different labels $\boldsymbol{y}_\tau^{(0)} = (y_\tau^{(0)1}, \ldots, y_\tau^{(0)\tau}) \in \mathcal{Y}^\tau$, $\boldsymbol{y}_\tau^{(1)} = (y_\tau^{(1)1}, \ldots, y_\tau^{(1)\tau}) \in \mathcal{Y}^\tau$ such that $y_\tau^{(0)i} \neq y_\tau^{(1)i}$ for each $i \in [\tau]$.
- Player $\mathtt{P}_L$ makes a sequence of τ binary choices $\boldsymbol{u}_\tau = (u_\tau^1, \ldots, u_\tau^\tau) \in \{0,1\}^\tau$ to produce the sequence of labels $\boldsymbol{y}_\tau = (y_\tau^1, \ldots, y_\tau^\tau) := \boldsymbol{r}(\boldsymbol{y}_\tau^{(0)}, \boldsymbol{y}_\tau^{(1)}, \boldsymbol{u}_\tau)$.

- Player $\mathtt{P}_L$ wins the game in round τ if $\mathcal{H}_{\boldsymbol{x}_1,\boldsymbol{y}_1,\ldots,\boldsymbol{x}_\tau,\boldsymbol{y}_\tau} = \varnothing$, where

$$\mathcal{H}_{\boldsymbol{x}_1,\boldsymbol{y}_1,\ldots,\boldsymbol{x}_\tau,\boldsymbol{y}_\tau} := \left\{h \in \mathcal{H} : h(x_s^i) = y_s^i \text{ for any } i \in [s], s \in [\tau]\right\}.$$

For convenience, we define the set $\widetilde{\mathcal{Y}}_t := \{(y^{(0)1}, \ldots, y^{(0)t}, y^{(1)1}, \ldots, y^{(1)t}) \in \mathcal{Y}^{2t} : y^{(0)i} \neq y^{(1)i},\ \forall i \in [t]\}$ which is the collection of all possible sequences of labels selected by $\mathtt{P}_A$ in round $t \in \mathbb{N}$. Then, in round $t \in \mathbb{N}$, the set of the plays of $\mathtt{P}_A$ is $\mathcal{X}^t \times \widetilde{\mathcal{Y}}_t$ and the set of plays of $\mathtt{P}_L$ is $\{0,1\}^t$. Define $\mathsf{P}_\infty := \prod_{t\in\mathbb{N}}(\mathcal{X}^t \times \widetilde{\mathcal{Y}}_t \times \{0,1\}^t)$. The set of winning sequences of $\mathtt{P}_L$ in $\mathscr{B}$ is

$$\mathsf{W}_{\mathscr{B}} := \left\{(\boldsymbol{x}_1, \widetilde{\boldsymbol{y}}_1, \boldsymbol{u}_1, \ldots) \in \mathsf{P}_\infty : \exists \tau \in \mathbb{N} \text{ such that } \mathcal{H}_{\boldsymbol{x}_1,\boldsymbol{r}(\widetilde{\boldsymbol{y}}_1,\boldsymbol{u}_1),\ldots,\boldsymbol{x}_\tau,\boldsymbol{r}(\widetilde{\boldsymbol{y}}_\tau,\boldsymbol{u}_\tau)} = \varnothing\right\}$$

As $\mathsf{W}_{\mathscr{B}}$ is finitely decidable, $\mathscr{B}$ is a Gale-Stewart game. Then, either $\mathtt{P}_A$ or $\mathtt{P}_L$ has a winning strategy (Gale and Stewart, 1953). Applying the proof of Kalavasis et al. (2022, Lemma 5) for countable $\mathcal{Y}$, we can see that $\mathtt{P}_A$ has a winning strategy in $\mathscr{B}$ if and only if $\mathcal{C}$ has an infinite Natarajan Littlestone tree, and if $\mathcal{C}$ does not have an infinite Natarajan Littlestone tree, then $\mathtt{P}_L$ has a universally measurable winning strategy. Let $\eta_t : \prod_{i=1}^{t-1}(\mathcal{X}^i \times \widetilde{\mathcal{Y}}_i) \times (\mathcal{X}^t \times \widetilde{\mathcal{Y}}_t) \to \{0,1\}^t$ denote this universally measurable winning strategy of $\mathtt{P}_L$ in round $t \in \mathbb{N}$.

Given an infinite sequence $((x_1, y_1), (x_2, y_2), \ldots) \in \mathcal{Z}^{\mathbb{N}}$, we define the following online algorithm.

1. Initialize $\tau_0 \leftarrow 1$ and $\boldsymbol{v}_0 \leftarrow \varnothing$.
2. For each round $t \in \mathbb{N}$:
   (a) Let $\tau_t \leftarrow \tau_{t-1}$, $\boldsymbol{v}_t \leftarrow \boldsymbol{v}_{t-1}$, $\boldsymbol{x} \leftarrow (x_{t-\tau_{t-1}+1}, \ldots, x_t)$, and $\boldsymbol{y} \leftarrow (y_{t-\tau_{t-1}+1}, \ldots, y_t)$.
   (b) For each $\widetilde{\boldsymbol{y}} \in \widetilde{\mathcal{Y}}_{\tau_{t-1}}$:
      i. Let $\boldsymbol{u} \leftarrow \eta_{\tau_{t-1}}(\boldsymbol{\xi}_{t-1}, \boldsymbol{x}, \widetilde{\boldsymbol{y}})$.
      ii. If $\boldsymbol{y} = \boldsymbol{r}(\widetilde{\boldsymbol{y}}, \boldsymbol{u})$:
         A. Let $\boldsymbol{\xi}_{\tau_{t-1}} \leftarrow (\boldsymbol{x}, \widetilde{\boldsymbol{y}})$, $\boldsymbol{v}_t \leftarrow (\boldsymbol{v}_{t-1}, \boldsymbol{x}, \widetilde{\boldsymbol{y}})$, and $\tau_t \leftarrow \tau_{t-1} + 1$.
         B. Break.

With the above algorithm, we can define the following mappings for each $t \in \mathbb{N}$,

$$T_{t-1} : \mathcal{Z}^{t-1} \to [t],\ ((x_1, y_1), \ldots, (x_{t-1}, y_{t-1})) \mapsto \tau_{t-1} \text{ and}$$
$$\boldsymbol{V}_{t-1} : \mathcal{Z}^{t-1} \to \cup_{i=0}^{t-1}\mathsf{P}_i,\ ((x_1, y_1), \ldots, (x_{t-1}, y_{t-1})) \mapsto \boldsymbol{v}_{t-1}.$$

Since η_t is universally measurable for any $t \in \mathbb{N}$, T_{t-1} and $\boldsymbol{V}_{t-1}$ are also universally measurable (Bousquet et al., 2021, Remark 5.4); (Hanneke et al., 2023, Proposition 52). Then, for any $t \in \mathbb{N} \cup \{0\}$ and $\boldsymbol{z} \in \mathcal{Z}^t$, we define

$$\boldsymbol{Y}_{\boldsymbol{z}} : \mathcal{X}^{T_t(\boldsymbol{z})} \times \to 2^{\mathcal{Y}^{T_t(\boldsymbol{z})}},\ \boldsymbol{x} \mapsto \{\boldsymbol{r}(\widetilde{\boldsymbol{y}}, \eta_{T_t(\boldsymbol{z})}(\boldsymbol{V}_{T_t(\boldsymbol{z})}(\boldsymbol{z}), \boldsymbol{x}, \widetilde{\boldsymbol{y}})) : \widetilde{\boldsymbol{y}} \in \widetilde{\mathcal{Y}}_{T_t(\boldsymbol{z})}\}.$$

Since η_t, T_{t-1} and $\boldsymbol{V}_{t-1}$ are universally measurable for all $t \in \mathbb{N}$, $\boldsymbol{Y}_{\boldsymbol{z}}$ is also universally measurable (Hanneke et al., 2023, Proposition 52). For any $k \in \mathbb{N}$ and universally measurable function $\boldsymbol{g} : \mathcal{X}^k \to 2^{\mathcal{Y}^k}$ where, we define

$$\mathrm{per}_P(\boldsymbol{g}) = \mathbb{P}_{((X_i,Y_i))_{i=1}^k \sim P^k}((Y_1, \ldots, Y_k) \in \boldsymbol{g}(X_1, \ldots, X_k))$$

for distribution P over $\mathcal{Z}$. According the proof of Kalavasis et al. (2022, Lemma 7), we have the following lemma.

Lemma 44 *Sample $\boldsymbol{Z} \sim P^{\mathbb{N}}$ and define $\boldsymbol{Y}_t := \boldsymbol{Y}_{\boldsymbol{Z}_{\leqslant t}}$ for $t \in \mathbb{N} \cup \{0\}$. Then, there exists a finite random variable $N \in \mathbb{N}$ such that for all $t \in \mathbb{N}$, $\mathrm{per}_P(\boldsymbol{Y}_t) > 0$ implies that $N > t$, almost surely.*

With bandit feedback, we also first "guess" the true labels using the distribution D over $\mathcal{Y}$ which has $D(y) > 0$ for all $y \in \mathcal{Y}$. Recall that we have defined the mapping $\hat{\boldsymbol{z}} : \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t) \to \mathcal{Z}^*$ such that $\hat{\boldsymbol{z}}(\varnothing) = \varnothing$ and for any $t \in \mathbb{N}$, $\boldsymbol{z} = ((x_1, y_1), \dots, (x_t, y_t)) \in \mathcal{Z}^t$, and $\boldsymbol{y} = (y_1', \dots, y_t') \in \mathcal{Y}^t$, $\hat{\boldsymbol{z}}(\boldsymbol{z}, \boldsymbol{y}) = ((x_i, y_i) : i \in [t], y_i = y_i')$. Now, for any $(\boldsymbol{z}, \boldsymbol{y}) \in \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t)$, we can define

$$\hat{\boldsymbol{Y}}_{\boldsymbol{z},\boldsymbol{y}} : \mathcal{X}^{T_{|\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})|}(\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y}))} \to 2^{\mathcal{Y}^{T_{|\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})|}(\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y}))}}, \ \boldsymbol{x} \mapsto \boldsymbol{Y}_{\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})}(\boldsymbol{x}).$$

Since $\boldsymbol{Y}_{\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})}$ is universally measurable, $\hat{\boldsymbol{Y}}_{\boldsymbol{z},\boldsymbol{y}}$ is also universally measurable. We can show the following result based on Lemma 44.

Lemma 45 *Sample $S = ((X_1, Y_1), (X_2, Y_2), \dots) \sim P^{\mathbb{N}}$ and $\boldsymbol{Y} = (Y_1', Y_2', \dots) \sim D^{\mathbb{N}}$ independently. For any $t \in \{0\} \cup \mathbb{N}$, define $\hat{\boldsymbol{Y}}_t := \hat{\boldsymbol{Y}}_{S_{\leqslant t}, \boldsymbol{Y}_{\leqslant t}}$. Then, we have*

$$\lim_{t\to\infty} \mathbb{P}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_t) > 0) = 0.$$

Proof According to Lemma 42, we have $|\boldsymbol{Z}| = \infty$, $\boldsymbol{Z} := \hat{\boldsymbol{z}}(S, \boldsymbol{Y}) \sim (P_D)^{\mathbb{N}}$ with $P_D \in \mathrm{RE}(\mathcal{C})$, and $K_t := |\hat{\boldsymbol{z}}(S_{\leqslant t}, \boldsymbol{Y}_{\leqslant t})| \to \infty$ as $t \to \infty$ almost surely. By Lemma 44, there exists $N \in \mathbb{N}$ finite such that $\mathrm{per}_{P_D}(\hat{\boldsymbol{Y}}_t) > 0$ implies that $N \geqslant K_t$ for all $t \in \mathbb{N}$ almost surely. Since $\lim_{t\to\infty} K_t = \infty$, we have $K_t > N$ for all sufficiently large t almost surely. It follows that $\lim_{t\to\infty} \mathbb{1}\{\mathrm{per}_{P_D}(\hat{\boldsymbol{Y}}_t) > 0\} = 0$ almost surely and $\lim_{t\to\infty} \mathbb{P}(\mathrm{per}_{P_D}(\hat{\boldsymbol{Y}}_t) > 0) = 0$.

For any $k \in \mathbb{N}$ and $g : \mathcal{X}^k \to 2^{\mathcal{Y}^k}$, define the set $\mathrm{PER}(g) := \{((x_i, y_i))_{i=1}^k \in \mathcal{Z}^k : (y_1, \dots, y_k) \in g(x_1, \dots, x_k)\}$. For any $\boldsymbol{y} = (y_1, \dots, y_k) \in \mathcal{Y}^k$, define the set $\mathcal{Z}_{\boldsymbol{y}} := \{((x_i, y_i))_{i=1}^k : x_i \in \mathcal{X}, \ \forall i \in [k]\}$. Then, if $\mathrm{per}_P(g) > 0$, since

$$\mathrm{per}_P(g) = \cup_{\boldsymbol{y}\in\mathcal{Y}^k} P^k(\mathrm{PER}(g) \cap \mathcal{Z}_{\boldsymbol{y}}) > 0,$$

there exists some $\boldsymbol{y}' = (y_1', \dots, y_k') \in \mathcal{Y}^k$ such that $P^k(\mathrm{PER}(g) \cap \mathcal{Z}_{\boldsymbol{y}'}) > 0$. Note that sampling $((X_i, Y_i))_{i=1}^k \sim P^k$ and $(Y_1', \dots, Y_k') \sim D^k$ independently, for any $E_1, \dots, E_k \subseteq \mathcal{Z}$ measurable, we have

$$\begin{aligned}
\mathbb{P}((X_i, Y_i) \in E_i, \ \forall i \in [k] \mid Y_i = Y_i', \ \forall i \in [k]) &= \frac{\mathbb{P}((X_i, Y_i) \in E_i, \ Y_i = Y_i', \ \forall i \in [k])}{\mathbb{P}(Y_i = Y_i', \ \forall i \in [k])} \\
&= \frac{\prod_{i=1}^k \mathbb{P}((X_i, Y_i) \in E_i, \ Y_i = Y_i')}{\prod_{i=1}^k \mathbb{P}(Y_i = Y_i')} \\
&= \prod_{i=1}^k P_D(E_i),
\end{aligned}$$

which implies that the conditional distribution of $((X_i, Y_i))_{i\in[k]} \mid Y_1 = Y_1', \dots, Y_k = Y_k'$ is $(P_D)^k$. Then, we have

$$\mathrm{per}_{P_D}(g) \geqslant (P_D)^k(\mathrm{PER}(g) \cap \mathcal{Z}_{\boldsymbol{y}'}) = \frac{\mathbb{P}((X_i, Y_i) \in \mathrm{PER}(g), \ Y_i = y_i', Y_i' = y_i', \ \forall i \in [k])}{(p_{P,D})^k}$$

$$= \frac{P^k(\mathrm{PER}(g) \cap \mathcal{Z}_{\boldsymbol{y}'}) \prod_{i=1}^k D(y_i')}{(p_{P,D})^k} > 0.$$

Since $\mathrm{per}_P(g) > 0$ implies that $\mathrm{per}_{P_D}(g) > 0$, we have

$$\mathbb{P}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_t) > 0) \leqslant \mathbb{P}(\mathrm{per}_{P_D}(\hat{\boldsymbol{Y}}_t) > 0), \ \forall t \in \mathbb{N},$$

and hence, $\lim_{t\to\infty} \mathbb{P}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_t) > 0) = 0$. ■

For any $k, n \in \mathbb{N}$ with $n \geqslant k$, $\boldsymbol{g} : \mathcal{X}^k \to 2^{\mathcal{Y}^k}$, and $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, we define the following concept class

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{g}) := \{h \in \mathcal{Y}^{[n]} : (h(i_1), \ldots, h(i_k)) \notin \boldsymbol{g}(x_{i_1}, \ldots, x_{i_k}), \ \forall \text{ pairwise distict } i_1, \ldots, i_k \in [n]\}.$$

For future reference, we show the following lemma.

Lemma 46 *For any permutation $\rho : [n] \to [n]$, we have $\mathcal{H}(\boldsymbol{x} \circ \rho, \boldsymbol{g}) = \mathcal{H}(\boldsymbol{x}, \boldsymbol{g}) \circ \rho$, where $\boldsymbol{x} \circ \rho := (x_{\rho(1)}, \ldots, x_{\rho(n)})$ and $\mathcal{H}(\boldsymbol{x}, \boldsymbol{g}) \circ \rho := \{h \circ \rho : h \in \mathcal{H}(\boldsymbol{x}, \boldsymbol{g})\}$.*

Proof For any $h \in \mathcal{H}(\boldsymbol{x}, \boldsymbol{g})$ and any pairwise distinct $i_1, \ldots, i_\tau \in [n]$, $\rho(i_1), \ldots, \rho(i_\tau) \in [n]$ are also pairwise distinct and we have

$$((h \circ \rho)(i_1), \ldots, (h \circ \rho)(i_\tau)) = (h(\rho(i_1)), \ldots, h(\rho(i_\tau))) \notin \boldsymbol{g}(x_{\rho(i_1)}, \ldots, x_{\rho(i_\tau)}),$$

which implies that $h \circ \rho \in \mathcal{H}(\boldsymbol{x} \circ \rho, \boldsymbol{g})$.

On the other hand, for any $h \in \mathcal{H}(\boldsymbol{x} \circ \rho, \boldsymbol{g})$, we define the function $f : [n] \to \mathcal{Y}, i \mapsto h(\rho^{-1}(i))$. Then, we have have $h = f \circ \rho$ and

$$(f(i_1), \ldots, f(i_\tau)) = (h(\rho^{-1}(i_1)), \ldots, h(\rho^{-1}(i_\tau))) \notin \boldsymbol{g}(x_{i_1}, \ldots, x_{i_\tau}),$$

which implies that $f \in \mathcal{H}(\boldsymbol{x}, \boldsymbol{g})$. ■

Generalizing (Kalavasis et al., 2022, Claim 2), we have the following result.

Lemma 47 *For any $t \in \mathbb{N} \cup \{0\}$ and $\boldsymbol{z} \in \mathcal{Z}^t$, let $\tau := T_t(\boldsymbol{z}) \in [t+1]$. For any $n \geqslant \tau$ and $\boldsymbol{x} \in \mathcal{X}^n$, we have $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})) < \tau$.*

Proof Write the sequence $\boldsymbol{x}$ as $\boldsymbol{x} = (x_1, \ldots, x_n)$. Suppose on the contrary that $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})) \geqslant \tau$. Then, there exists $\boldsymbol{i} = (i_1, \ldots, i_\tau) \in [n]^\tau$ with $i_1 < \cdots < i_\tau$ and

$$\widetilde{\boldsymbol{y}} = (y^{(0)1}, \ldots, y^{(0)\tau}, y^{(1)1}, \ldots, y^{(1)\tau}) \in \widetilde{\mathcal{Y}}_\tau$$

such that $\prod_{i=1}^\tau \{y^{(0)i}, y^{(1)i}\} \subseteq \mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})|_{\boldsymbol{i}}$. However, by the definition of $\mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})$, we have

$$(h(i_1), \ldots, h(i_\tau)) \notin \boldsymbol{Y}_{\boldsymbol{z}}(x_{i_1}, \ldots, x_{i_\tau})$$

for any $h \in \mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})$. By the definition of $\boldsymbol{Y}_{\boldsymbol{z}}$, there exists some $\boldsymbol{u} \in \{0,1\}^\tau$ such that $\boldsymbol{r}(\widetilde{\boldsymbol{y}}, \boldsymbol{u}) \in \boldsymbol{Y}_{\boldsymbol{z}}(x_{i_1}, \ldots, x_{i_\tau})$. Then, we have $\boldsymbol{r}(\widetilde{\boldsymbol{y}}, \boldsymbol{u}) \in \prod_{i=1}^\tau \{y^{(0)i}, y^{(1)i}\}$ and $\boldsymbol{r}(\widetilde{\boldsymbol{y}}, \boldsymbol{u}) \notin \mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})|_{\boldsymbol{i}}$, which contradicts $\prod_{i=1}^\tau \{y^{(0)i}, y^{(1)i}\} \subseteq \mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})|_{\boldsymbol{i}}$. Thus, we have $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{z}})) < \tau$. ■

The next corollary follows directly from the above lemma.

Corollary 48 *For any $t \in \mathbb{N} \cup \{0\}$, $\boldsymbol{y} \in \mathcal{Y}^t$, and $\boldsymbol{z} \in \mathcal{Z}^t$, let $\tau := T_{|\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})|}(\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})) \in [t+1]$. For any $n \geqslant \tau$ and $\boldsymbol{x} \in \mathcal{X}^n$, we have $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{Y}}_{\boldsymbol{z},\boldsymbol{y}})) < \tau$.*

Finally, for notational convenience, we introduce the mapping

$$\hat{\tau} : \cup_{t=0}^{\infty}(\mathcal{Z}^t \times \mathcal{Y}^t) \to \mathbb{N},\ (\boldsymbol{z},\boldsymbol{y}) \mapsto T_{|\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})|}(\hat{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{y})), \tag{4}$$

where $\boldsymbol{z} \in \mathcal{Z}^t$ and $\boldsymbol{y} \in \mathcal{Y}^t$ for some $t \in \mathbb{N} \cup \{0\}$. Since T_{t-1} is universally measurable for any $t \in \mathbb{N}$ and $\hat{\boldsymbol{z}}$ is measurable, $\hat{\tau}$ is universally measurable.

E.3. Applying the One-Inclusion Algorithm with Constructed Concept Class

We first introduce the one-inclusion algorithm (Brukhim et al., 2022, Algorithm 1) which is the building block of our learning algorithm. Let $\mathscr{X}$ be an arbitrary instance set and $\mathscr{Y}$ be an arbitrary label set. For an arbitrary concept class $\mathcal{F} \subseteq \mathscr{Y}^{\mathscr{X}}$, the one-inclusion algorithm $\mathcal{A}_{\mathcal{F}} : (\mathscr{X} \times \mathscr{Y})^* \to \mathscr{Y}^{\mathscr{X}}$ is defined as follows.

Algorithm 1: The one-inclusion algorithm $\mathcal{A}_{\mathcal{F}}$ (Brukhim et al., 2022, Algorithm 1)

Input: A $\mathcal{F}$-realizable sequence $\boldsymbol{s} = ((x_1, y_1), \ldots, (x_n, y_n)) \in (\mathscr{X} \times \mathscr{Y})$ with $n \in \mathbb{N}$.
Output: A classifier $\mathcal{A}_{\mathcal{F}}(\boldsymbol{s}) = h_{\boldsymbol{s}} \in \mathscr{Y}^{\mathscr{X}}$.
Given $x \in \mathcal{X}$, the value of $h_{\boldsymbol{s}}(x) \in \mathscr{Y}$ is calculated as follows:

1. Consider $\mathcal{F}|_{(x_1,\ldots,x_n,x)} \subseteq \mathscr{Y}^{n+1}$.
2. Find an orientation σ of $\mathsf{OIG}(\mathcal{F}|_{(x_1,\ldots,x_n,x)})$ that minimizes the maximum out-degree.
3. Let $e \leftarrow \{h \in \mathcal{F}|_{(x_1,\ldots,x_n,x)} : h(i) = y_i,\ \forall i \in [n]\}$.
4. Let $h_{\boldsymbol{s}}(x) \leftarrow \sigma((e, n+1))(n+1)$.

For the above algorithm, we provide the definitions of OIG (one-inclusion graph), orientation, and maximum out-degree below.

Definition 49 (One-inclusion graph OIG, Haussler et al. 1994) *The **one-inclusion graph** (OIG) of $H \subseteq \mathscr{Y}^n$ for $n \in \mathbb{N}$ is a hypergraph $\mathsf{OIG}(H) = (H, E)$ where H is the vertex-set and E denotes the edge-set defined as follows. For any $i \in [n]$ and $f : [n]\backslash\{i\} \to \mathcal{Y}$, we define the set $e_{i,f} := \{h \in H : h(j) = f(j),\ \forall j \in [n]\backslash\{i\}\}$. Then, the edge-set is defined as*

$$E := \{(e_{i,f}, i) : i \in [n], f : [n]\backslash\{i\} \to \mathcal{Y}, e_{i,f} \neq \varnothing\}.$$

For any $(e_{i,f}, i) \in E$ and $h \in H$, we say $h \in (e_{i,f}, i)$ if $h \in e_{i,f}$ and the size of the edge is $|(e_{i,f}, i)| := |e_{i,f}|$.

Definition 50 (Orientation, Brukhim et al. 2022, Definition 11) *An **orientation** of a hypergraph (V, E) is a mapping $\sigma : E \to V$ such that $\sigma(e) \in e$ for each edge $e \in E$.*

Definition 51 (Out-degree) *Given a hypergraph (V, E) and an orientation $\sigma : E \to V$, the **out-degree** of a vertex $v \in V$ is $\mathsf{outdeg}(v; \sigma) := |\{e \in E : v \in e \text{ and } \sigma(e) \neq v\}|$ and the **maximum out-degree** of σ is $\mathsf{outdeg}(\sigma) := \sup_{v \in V} \mathsf{outdeg}(v; \sigma)$.*

We have the following guarantee on Algorithm 1.

Lemma 52 *Let $n \in \mathbb{N}$, $\mathcal{F} \subseteq \mathcal{Y}^{[n+1]}$ with $\mathrm{ND}(\mathcal{F}) \leqslant d \in [n]$, and $(y_1, \ldots, y_{n+1}) \in \mathcal{Y}^{n+1}$. Define $f_* : [n+1] \to \mathcal{Y}$, $i \mapsto y_i$. Suppose that $|\{f(i) : f \in \mathcal{F}\}| \leqslant k \in \mathbb{N}$ for all $i \in [n+1]$ and $f_* \in \mathcal{F}$. Then, the one-inclusion algorithm $\mathcal{A}_{\mathcal{F}}$ defined by Algorithm 1 satisfies that*

$$\sum_{i=1}^{n+1} \mathbb{1}\{\mathcal{A}_{\mathcal{F}}(\boldsymbol{s}_{-i}, i) \neq y_i\} \leqslant 20d \log(k),$$

where $\boldsymbol{s} = ((j, y_j))_{j \in [n+1]}$ and $\boldsymbol{s}_{-i} = ((j, y_j))_{j \in [n+1] \setminus \{i\}}$ for any $i \in [n+1]$.

Proof Since $\mathcal{Y}$ is countable, we can let $\{y^1, y^2, \ldots\}$ be an enumeration of $\mathcal{Y}$. Then, for any subset $\mathsf{Y} \subseteq \mathcal{Y}$, we can enumerate it as $\mathsf{Y} = \{y^{i_1}, y^{i_2}, \cdots\}$ such that $1 \leqslant i_1 < i_2 < \cdots$. Now, we can define the mapping $\phi_{\mathsf{Y}} : \mathsf{Y} \to [k]$ such that $\phi_{\mathsf{Y}}(y^{i_j}) = j$ if $j \in [k]$ and $\phi_{\mathsf{Y}}(y^{i_j}) = k$ if $j > k$. Conversely, we can define the mapping $\psi_{\mathsf{Y}} : [k] \to \mathsf{Y}$ such that $\psi_{\mathsf{Y}}(j) = y^{i_j}$ for all $j \in [k]$.

Now, defining $\mathcal{F}_i := \{f(i) : f \in \mathcal{F}\} \subseteq \mathcal{Y}$ for all $i \in [n+1]$, we can define

$$f_{\mathcal{F}} : [n+1] \to [k],\ i \mapsto \phi_{\mathcal{F}_i}(f(i))$$

for any concept $f \in \mathcal{F}$. Define $g_* = (f_*)_{\mathcal{F}}$ and consider the following concept class

$$\mathcal{G} := \{f_{\mathcal{F}} : f \in \mathcal{F}\} \subseteq [k]^{[n+1]}.$$

Since $f_* \in \mathcal{F}$, we have $g_* \in \mathcal{G}$. Since $|\mathcal{F}_i| \leqslant k$ for all $i \in [n+1]$, there is a bijective mapping $\alpha : \mathcal{F} \to \mathcal{G}$, $f \mapsto f_{\mathcal{F}}$. If there exist $1 \leqslant i_1 < \cdots < i_{d+1} \leqslant n+1$ and everywhere different $g_1, g_2 : [d+1] \to [k]$ such that $\prod_{j=1}^{d+1} \{g_1(j), g_2(j)\} \subseteq \mathcal{G}|_{(i_1, \ldots, i_{d+1})}$, then we can define $f_\ell : [d+1] \to \mathcal{Y}$, $j \mapsto \psi_{\mathcal{F}_{i_j}}(g_\ell(j))$ for $\ell = 1, 2$ and have $\prod_{j=1}^{d+1} \{f_1(j), f_2(j)\} \subseteq \mathcal{F}|_{(i_1, \ldots, i_{d+1})}$, which proves that $\mathrm{ND}(\mathcal{G}) \leqslant d$.

Define $\boldsymbol{r} := ((j, g_*(j)))_{j \in [n+1]}$. By the definition of the one-inclusion algorithm (Brukhim et al., 2022, Algorithm 1), we have $\mathcal{A}_{\mathcal{G}}(\boldsymbol{r}_{-i}, i) = \phi_{\mathcal{F}_i}(\mathcal{A}_{\mathcal{F}}(\boldsymbol{s}_{-i}, i))$ for any $i \in [n+1]$. By Brukhim et al. (2022, Lemma 17), since $g_* \in \mathcal{G}$, we have

$$\sum_{i=1}^{n+1} \mathbb{1}\{\mathcal{A}_{\mathcal{F}}(\boldsymbol{s}_{-i}, i) \neq y_i\} = \sum_{i=1}^{n+1} \mathbb{1}\{\mathcal{A}_{\mathcal{G}}(\boldsymbol{r}_{-i}, i) \neq \phi_{\mathcal{F}_i}(y_i)\} \leqslant 20d \log(k).$$

■

Suppose that we have constructed some menu $\mu : \mathcal{X} \to \mathcal{Y}$ with $|\mu| \leqslant k \in \mathbb{N}$ and universally measurable function $\boldsymbol{g} : \mathcal{X}^\tau \to 2^{\mathcal{Y}^\tau}$ with $\tau \in \mathbb{N}$. Then, for any $n \in \mathbb{N}$, $x \in \mathcal{X}$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, and $\mu : \mathcal{X} \to 2^{\mathcal{Y}}$, we define $\boldsymbol{s} := ((x_i, y_i))_{i=1}^n$ and if $n \geqslant \tau$,

$$\mathcal{F}_{(\boldsymbol{x}, x), \boldsymbol{g}, \mu} := \{h \in \mathcal{H}((\boldsymbol{x}, x), \boldsymbol{g}) : h(j) \in \mu(x_j), \forall j \in [n] \text{ and } h(n+1) \in \mu(x)\} \subseteq \mathcal{Y}^{[n+1]}.$$

Now, fixing an arbitrary $y_0 \in \mathcal{Y}$ and defining $\widetilde{\boldsymbol{s}} := ((i, y_i))_{i=1}^n$, we can construct the algorithm $\mathbb{A}_{\boldsymbol{g},\mu} : (\mathcal{X} \times \mathcal{Y})^* \times \mathcal{X} \to \mathcal{Y}$ by defining

$$\mathbb{A}_{\boldsymbol{g},\mu}(\boldsymbol{s}, x) := \begin{cases} \mathcal{A}_{\mathcal{F}_{(\boldsymbol{x},x),\boldsymbol{g},\mu}}(\widetilde{\boldsymbol{s}}, n+1), \text{ if } n \geqslant \tau \text{ and } \widetilde{\boldsymbol{s}} \text{ is } \mathcal{F}_{(\boldsymbol{x},x),\boldsymbol{g},\mu}\text{-realizable,} \\ y_0, \text{ otherwise.} \end{cases} \tag{5}$$

We upper bound the expected error rate of $\mathbb{A}_{\boldsymbol{g},\mu}$ in the following lemma.

Lemma 53 *Assume that $|\mu| \leqslant k \in \mathbb{N}$, $\mathrm{er}_P(\mu) = 0$, $\mathrm{per}_P(\boldsymbol{g}) = 0$, and $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \boldsymbol{g})) \leqslant d$ for all $\boldsymbol{x} \in \mathcal{X}^n$ with $n \geqslant \tau$. Then, for any $n \in \mathbb{N}$ with $n \geqslant \tau$, we have*

$$\mathbb{P}_{(S,(X,Y))\in P^n}(\mathbb{A}_{\boldsymbol{g},\mu}(S, X) \neq Y) \leqslant \frac{20d\log(k)}{n+1}.$$

Proof Sample $\bar{S} = ((X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})) \sim P^{n+1}$. Since $\mathrm{per}_P(\boldsymbol{g}) = 0$, we have $(Y_{i_1}, \ldots, Y_{i_\tau}) \notin \boldsymbol{g}(X_{i_1}, \ldots, X_{i_\tau})$ for all $1 \leqslant i_1 < \cdots < i_\tau \leqslant n+1$ almost surely. It follows that the function $f_* : [n+1] \to \mathcal{Y}, i \mapsto Y_i$ is contained in $\mathcal{H}(\boldsymbol{X}, \boldsymbol{g})$, where $\boldsymbol{X} := (X_1, \ldots, X_{n+1})$. Since $\mathrm{er}_P(\mu) = 0$, we have $f_*(i) = Y_i \in \mu(X_i)$ for all $i \in [n+1]$ almost surely, which implies that $f_* \in \mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}$. Moreover, by assumption, we have $\mathrm{ND}(\mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}) \leqslant \mathrm{ND}(\mathcal{H}(\boldsymbol{X}, \boldsymbol{g})) \leqslant d$ and $|\{f(i) : f \in \mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}\}| \leqslant |\mu| \leqslant k$ for all $i \in [n+1]$.

Now, we consider the permutation $\rho_i : [n] \to [n], j \mapsto j\mathbb{1}\{j < i\} + (j+1)\mathbb{1}\{i \leqslant j \leqslant n\} + i\mathbb{1}\{j = n+1\}$ for each $i \in [n+1]$. Then, defining $\widetilde{S}^i := ((j, Y_{\rho_i(j)}))_{j=1}^n$, we have

$$\mathbb{A}_{\boldsymbol{g},\mu}(\bar{S}_{-i}, X_i) = \mathcal{A}_{\mathcal{F}_{(\boldsymbol{X}_{-i},X_i),\boldsymbol{g},\mu}}(\widetilde{S}^i, n+1).$$

Since $\mathcal{H}((\boldsymbol{X}_{-i}, X_i), \boldsymbol{g}) = \mathcal{H}(\boldsymbol{X}, \boldsymbol{g}) \circ \rho_i$ by Lemma 46, we have $\mathcal{F}_{(\boldsymbol{X}_{-i},X_i),\boldsymbol{g},\mu} = \mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu} \circ \rho_i$ and

$$\mathbb{A}_{\boldsymbol{g},\mu}(\bar{S}_{-i}, X_i) = \mathcal{A}_{\mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}\circ\rho_i}(\widetilde{S}^i, n+1) = \mathcal{A}_{\mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}}(\widetilde{S}_{-i}, i),$$

where $\widetilde{S} := ((j, Y_j))_{j=1}^{n+1}$. Furthermore, by Lemma 52, we have

$$\sum_{i=1}^{n+1} \mathbb{1}\{\mathbb{A}_{\boldsymbol{g},\mu}(\bar{S}_{-i}, X_i) \neq Y_i\} = \sum_{i=1}^{n+1} \mathbb{1}\{\mathcal{A}_{\mathcal{F}_{\boldsymbol{X},\boldsymbol{g},\mu}}(\widetilde{S}_{-i}, i) \neq Y_i\} \leqslant 20d\log(k)$$

almost surely. Since $\mathbb{P}_{(S,(X,Y))\in P^n}(\mathbb{A}_{\boldsymbol{g},\mu}(S, X) \neq Y) = \mathbb{E}[\mathbb{1}\{\mathbb{A}_{\boldsymbol{g},\mu}(\bar{S}_{-i}, X_i) \neq Y_i\}]$ for all $i \in [n+1]$, we have

$$\mathbb{P}_{(S,(X,Y))\in P^n}(\mathbb{A}_{\boldsymbol{g},\mu}(S, X) \neq Y) = \frac{\mathbb{E}\big[\sum_{i=1}^{n+1} \mathbb{1}\{\mathbb{A}_{\boldsymbol{g},\mu}(\bar{S}_{-i}, X_i) \neq Y_i\}\big]}{n+1} \leqslant \frac{20d\log(k)}{n+1}.$$

■

E.4. Learning Algorithm and Its Expected Error Rate

Now, we are ready to prove Theorem 33.

Proof [Proof of Theorem 33] By Lemma 43 and Lemma 45, there exists some $t_* \in \mathbb{N}$ depending on P such that for all $t \geqslant t_*$, we have

$$\mathbb{P}_{(\boldsymbol{Z},\boldsymbol{Y})\sim P^{\mathbb{N}}\times D^{\mathbb{N}}}(\exists j \geqslant t_* \text{ s.t. } \mathrm{er}_P(\hat{\mu}_{\boldsymbol{Z}_{\leqslant j},\boldsymbol{Y}_{\leqslant j}}) > 0 \text{ or } |\hat{\mu}_{\boldsymbol{Z}_{\leqslant j},\boldsymbol{Y}_{\leqslant j}}| > t_*) < 1/8 \text{ and}$$
$$\mathbb{P}_{(\boldsymbol{Z},\boldsymbol{Y})\sim P^{\mathbb{N}}\times D^{\mathbb{N}}}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_{\boldsymbol{Z}_{\leqslant t},\boldsymbol{Y}_{\leqslant t}})) < 1/8.$$

Then, we also have

$$\begin{aligned}
&\mathbb{P}_{(\boldsymbol{Z}_t,\boldsymbol{Y}_t)\sim P^t\times D^t}(\mathrm{er}_P(\hat{\mu}_{\boldsymbol{Z}_t,\boldsymbol{Y}_t}) > 0 \text{ or } |\hat{\mu}_{\boldsymbol{Z}_t,\boldsymbol{Y}_t}| > t_*)\\
=&\mathbb{P}_{(\boldsymbol{Z},\boldsymbol{Y})\sim P^{\mathbb{N}}\times D^{\mathbb{N}}}(\mathrm{er}_P(\hat{\mu}_{\boldsymbol{Z}_{\leqslant t},\boldsymbol{Y}_{\leqslant t}}) > 0 \text{ or } |\hat{\mu}_{\boldsymbol{Z}_{\leqslant t},\boldsymbol{Y}_{\leqslant t}}| > t_*)\\
\leqslant&\mathbb{P}_{(\boldsymbol{Z},\boldsymbol{Y})\sim P^{\mathbb{N}}\times D^{\mathbb{N}}}(\exists j \geqslant t_* \text{ s.t. } \mathrm{er}_P(\hat{\mu}_{\boldsymbol{Z}_{\leqslant j},\boldsymbol{Y}_{\leqslant j}}) > 0 \text{ or } |\hat{\mu}_{\boldsymbol{Z}_{\leqslant j},\boldsymbol{Y}_{\leqslant j}}| > t_*) < 1/8.
\end{aligned}$$

and

$$\mathbb{P}_{(\boldsymbol{Z}_t,\boldsymbol{Y}_t)\sim P^t\times D^t}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_{\boldsymbol{Z}_t,\boldsymbol{Y}_t}) > 0) = \mathbb{P}_{(\boldsymbol{Z},\boldsymbol{Y})\sim P^{\mathbb{N}}\times D^{\mathbb{N}}}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_{\boldsymbol{Z}_{\leqslant t},\boldsymbol{Y}_{\leqslant t}})) < 1/8.$$

For training sample size $n \in \mathbb{N}$, sample training data $S = ((X_1, Y_1), \ldots, (X_n, Y_n)) \sim P^n$. Define $m := \lfloor\sqrt{n/2}\rfloor$. Sample $\boldsymbol{Y}' = (Y'_1, \ldots, Y'_{\lfloor n/2\rfloor}) \sim D^{\lfloor n/2\rfloor}$. For each $i \in [m]$, define the subsequences $S^{(i)} := ((X_j, Y_j))_{j=(i-1)m+1}^{im}$ and $\boldsymbol{Y}'^{(i)} := (Y'_j)_{j=(i-1)m+1}^{im}$. Then, we can define the menu

$$\mu_i : \mathcal{X} \to 2^{\mathcal{Y}},\ x \mapsto \hat{\boldsymbol{L}}(S^{(i)}, \boldsymbol{Y}'^{(i)}, x).$$

Note that $\mu_1, \ldots, \mu_m$ are independent. If $m \geqslant t_*$, we have

$$\mathbb{P}(\mathrm{er}_P(\mu_i) > 0 \text{ or } |\mu_i| > t_*) < 1/8$$

for all $i \in [m]$. Then, by Hoeffding's inequality, we have

$$\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\{\mathrm{er}_P(\mu_i) > 0 \text{ or } |\mu_i| > t_*\} \geqslant \frac{1}{4}\right) \leqslant e^{-m/32}. \tag{6}$$

Define the index set $\mathcal{I} := \{i \in [m] : \mathrm{er}_P(\mu_i) = 0 \text{ and } |\mu_i| \leqslant t_*\}$. Then, (6) implies that $|\mathcal{I}| > \frac{3m}{4}$ with probability at least $1 - e^{-m/32}$. Define the menu

$$\hat{\mu} : \mathcal{X} \to 2^{\mathcal{Y}},\ x \mapsto \{y \in \mathcal{Y} : |\{i \in [m] : y \in \mu_i(x)\}| > 3m/4\}.$$

Note that $\hat{\mu}$ is constructed only using $S_{\leqslant m^2}$ with $m^2 \leqslant \lfloor n/2\rfloor$. Under the event that $|\mathcal{I}| > \frac{3m}{4}$, we have

$$\begin{aligned}
1 - \mathrm{er}_P(\hat{\mu}) =&P(\{(x, y) \in \mathcal{Z} : |\{i \in [m] : y \in \mu_i(x)\}| > 3m/4\})\\
\geqslant&P(\{(x, y) \in \mathcal{Z} : y \in \mu_i(x),\ \forall i \in \mathcal{I}\})
\end{aligned}$$

$$\begin{aligned}\geqslant & 1-\sum_{i\in\mathcal{I}} P(\{(x,y)\in\mathcal{Z}: y\notin\mu_i(x)\}) \\ = & 1-\sum_{i\in\mathcal{I}} \mathrm{er}_P(\mu_i) = 1,\end{aligned}$$

i.e., $\mathrm{er}_P(\hat{\mu})=0$. For any $(x,y)\in\mathcal{Z}$, under the even that $|\mathcal{I}|>3m/4$, if $y\in\hat{\mu}(x)$, we must have

$$\{i\in\mathcal{I}: y\in\mu_i(x)\} > 3m/4 - |[m]\backslash\mathcal{I}| > m/2$$

and therefore,

$$\frac{m}{2}|\hat{\mu}(x)| = \sum_{y\in\mathcal{Y}}\frac{m}{2}\mathbb{1}\{y\in\hat{\mu}(x)\} < \sum_{y\in\mathcal{Y}}\sum_{i\in\mathcal{I}}\mathbb{1}\{y\in\mu_i(x)\} \leqslant \sum_{i\in\mathcal{I}}|\mu_i| \leqslant mt_*,$$

which implies that $|\hat{\mu}|<2t_*$. Thus, by (6), we have that $|\hat{\mu}|\leqslant 2t_*$ and $\mathrm{er}_P(\hat{\mu})=0$ with probability at least $1-e^{-m/32}$.

By Hanneke et al. (2023, Lemma 62), there exists some constant $M\geqslant 1$ such that if $n\geqslant \max\{4(t_*+1), M\}$, then there exists a universally measurable function $\hat{T}_n: \mathcal{Z}^{\lfloor n/2\rfloor}\to[\lfloor n/4\rfloor-1]$ such that $\hat{t}_n=\hat{T}_n(X_1,Y_1,\ldots,X_{\lfloor n/2\rfloor},Y_{\lfloor n/2\rfloor})$ such that

$$\mathbb{P}(\hat{t}_n\in\mathfrak{T}_{\mathrm{good}})\geqslant 1-Ce^{-cn}$$

for some constants $c,C>0$, where $\mathfrak{T}_{\mathrm{good}}:=\{t\in[t_*]:\mathbb{P}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_{S_{\leqslant t},\boldsymbol{Y}'_{\leqslant t}})>0)\leqslant 3/8\}$.

Now, defining the event $\mathcal{E}:=\{\mathrm{er}_P(\hat{\mu})=0 \text{ and } |\hat{\mu}|\leqslant 2t_* \text{ and } \hat{t}_n\in\mathfrak{T}_{\mathrm{good}}\}$, by union bound and the above results, we have

$$\mathbb{P}(\mathcal{E}^c)\leqslant e^{-m/32}+Ce^{-cn}.$$

Next, we define $\hat{m}:=\lfloor n/(2\hat{t}_n)\rfloor$. For any $t\in[\lfloor n/4\rfloor-1]$ and $i\in\lfloor n/(2t)\rfloor$, define the subsequences $S^{(t,i)}:=((X_j,Y_j))_{j=(i-1)t+1}^{it}$ and $\boldsymbol{Y}'^{(t,i)}:=((Y'_j))_{j=(i-1)t+1}^{it}$. If $t\in\mathfrak{T}_{\mathrm{good}}$, we have $\mathbb{P}(\mathrm{per}_P(\hat{\boldsymbol{Y}}_{S^{(t,i)},\boldsymbol{Y}'^{(t,i)}})>0)\leqslant 3/8$ and $t\leqslant t_*$. Then, by Hoeffding's inequality, we have

$$\mathbb{P}\left(\frac{1}{\lfloor n/(2t)\rfloor}\sum_{i=1}^{\lfloor n/(2t)\rfloor}\mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}_{S^{(t,i)},\boldsymbol{Y}'^{(t,i)}})>0\}\geqslant\frac{7}{16}\right)\leqslant e^{-\lfloor n/(2t)\rfloor/128}\leqslant e^{-\lfloor n/(2t_*)\rfloor/128}.$$

Define the mappings

$$\hat{\boldsymbol{Y}}^i:=\hat{\boldsymbol{Y}}_{S^{(\hat{t}_n,i)},\boldsymbol{Y}'^{(\hat{t}_n,i)}},\ \forall i\in[\hat{m}].$$

It follows from union bound that

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{\hat{m}}\sum_{i=1}^{\hat{m}}\mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i)>0\}\geqslant\frac{7}{16},\ \hat{t}_n\in\mathfrak{T}_{\mathrm{good}}\right)\\ \leqslant&\sum_{t\in\mathfrak{T}_{\mathrm{good}}}\mathbb{P}\left(\frac{1}{\hat{m}}\sum_{i=1}^{\hat{m}}\mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i)>0\}\geqslant\frac{7}{16}\right)\leqslant t_*e^{-\lfloor n/(2t_*)\rfloor/128}.\end{aligned}$$

Define $n' := n - \lfloor n/2 \rfloor$ and $\widetilde{S} := ((X_{\lfloor n/2 \rfloor + i}, Y_{\lfloor n/2 \rfloor + i}))_{i=1}^{n'}$. Define $\tau_i := \hat{\tau}(S^{(\hat{t}_n, i)}, \boldsymbol{Y}'^{(\hat{t}_n, i)})$ (see (4) for the definition of $\hat{\tau}$) for each $i \in [\hat{m}]$. Note $\tau_i \leqslant \hat{t}_n + 1 \leqslant \lfloor n/4 \rfloor < n'$. For any $i \in [\hat{m}]$, we define the classifier $\hat{h}_i : \mathcal{X} \to \mathcal{Y}$ such that

$$\hat{h}_i(x) := \mathbb{A}_{\hat{\boldsymbol{Y}}^i, \hat{\mu}}(\widetilde{S}, x),$$

where $\mathbb{A}_{\hat{\boldsymbol{Y}}^i, \hat{\mu}}$ is defined by (5). By Corollary 48, we have $\mathrm{ND}(\mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{Y}}^i)) \leqslant \tau_i - 1 \leqslant \hat{t}_n \leqslant t_*$ for all $\boldsymbol{x} \in \mathcal{X}^\ell$ with $\ell \in \mathbb{N}$ and $\ell \geqslant \tau_i$ under the event $\mathcal{E}$. Also note that $\hat{\mu}$, $\hat{t}_n$, and $\hat{\boldsymbol{Y}}^1, \ldots, \hat{\boldsymbol{Y}}^{\hat{m}}$ are determined by $((X_j, Y_j))_{j=1}^{\lfloor n/2 \rfloor}$ which is independent of $\widetilde{S}$, and $\widetilde{S} \sim P^{n'}$. Then, sampling $(X, Y) \sim P$ independent of the training sample S, by Lemma 53, we have

$$\mathbb{1}_{\mathcal{E}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) = 0\} \mathbb{E}\Big[\mathbb{1}\{\hat{h}_i(X) \neq Y\} \mid ((X_j, Y_j))_{j=1}^{\lfloor n/2 \rfloor}\Big] \leqslant \frac{20 t_* \log(2t_*)}{n' + 1} \leqslant \frac{40 t_* \log(2t_*)}{n},$$

which yields that

$$\mathbb{E}\Big[\mathbb{1}_{\mathcal{E}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) = 0\} \mathbb{1}\{\hat{h}_i(X) \neq Y\} \mid ((X_j, Y_j))_{j=1}^{\lfloor n/2 \rfloor}\Big] \leqslant \frac{40 t_* \log(2t_*)}{n}.$$

Our final output classifier is $\hat{h}_S := \mathrm{Maj}(\hat{h}_1, \ldots, \hat{h}_{\hat{m}})$, the majority vote of the $\hat{m}$ classifiers $\hat{h}_1, \ldots, \hat{h}_{\hat{m}}$. Given the above results, we can apply Markov's inequality to have

$$\begin{aligned}
\mathbb{E}[\mathrm{er}_P(\hat{h}_S)] =& \mathbb{P}(\hat{h}_S(X) \neq Y) \\
\leqslant& \mathbb{P}(\mathcal{E}^c) + \mathbb{P}\left(\mathcal{E},\ \frac{1}{\hat{m}} \sum_{i=1}^{\hat{m}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) > 0\} \geqslant \frac{7}{16}\right) \\
&+ \mathbb{P}\left(\mathcal{E},\ \frac{1}{\hat{m}} \sum_{i=1}^{\hat{m}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) = 0\} > \frac{9}{16},\ \frac{1}{\hat{m}} \sum_{i=1}^{m} \mathbb{1}\{\hat{h}_i(X) \neq Y\} \geqslant \frac{1}{2}\right) \\
\leqslant& e^{-m/32} + Ce^{-cn} + t_* e^{-\lfloor n/(2t_*) \rfloor / 128} \\
&+ \mathbb{P}\left(\frac{1}{\hat{m}} \sum_{i=1}^{\hat{m}} \mathbb{1}_{\mathcal{E}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) = 0\} \mathbb{1}\{\hat{h}_i(X) \neq Y\} > \frac{1}{16}\right) \\
\leqslant& e^{-m/32} + Ce^{-cn} + t_* e^{-\lfloor n/(2t_*) \rfloor / 128} \\
&+ 16\mathbb{E}\left[\frac{1}{\hat{m}} \sum_{i=1}^{\hat{m}} \mathbb{E}\Big[\mathbb{1}_{\mathcal{E}} \mathbb{1}\{\mathrm{per}_P(\hat{\boldsymbol{Y}}^i) = 0\} \mathbb{1}\{\hat{h}_i(X) \neq Y\} \mid ((X_j, Y_j))_{j=1}^{\lfloor n/2 \rfloor}\Big]\right] \\
\leqslant& e^{-\lfloor \sqrt{n/2} \rfloor / 32} + Ce^{-cn} + t_* e^{-\lfloor n/(2t_*) \rfloor / 128} + \frac{640 t_* \log(2t_*)}{n}.
\end{aligned}$$

Thus, $\mathcal{Q}$ is universal learnable under bandit feedback at rate $1/n$. ■

Appendix F. Universal Consistency

Theorem 54 *$\mathcal{Q} = (\mathcal{X}, \mathcal{Y}, \mathcal{C})$ is universal learnable under bandit feedback.*

Proof Since $\mathcal{X}$ is Polish and $\mathcal{Y}$ is countable, there exists a deterministic learning algorithm $\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{Y}^{\mathcal{X}}$ under full-information feedback such that $\lim_{n\to\infty} \mathbb{E}_{S\sim P^n}[\mathrm{er}_P(\mathcal{A}(S))] = 0$ for all realizable distributions $P \in \mathrm{RE}(\mathcal{C})$ (Hanneke et al., 2021).

As in Section E, let $D \in \Pi(\mathcal{Y})$ be a probability measure on $\mathcal{Y}$ such that $D(y) := D(\{y\}) > 0$ for all $y \in \mathcal{Y}$. Define the mapping $\hat{\boldsymbol{z}} : (\mathcal{X} \times \Sigma)^* \to \mathcal{Z}^*$ (recall that $\Sigma = \mathcal{Y} \times \{0, 1\}$) such that $\hat{\boldsymbol{z}}(\varnothing) := \varnothing$ and for any $t \in \mathbb{N}$ and $\boldsymbol{\xi} = ((x_1, y_1, b_1), \ldots, (x_t, y_t, b_t)) \in (\mathcal{X} \times \Sigma)^t$, let $\hat{\boldsymbol{z}}(\boldsymbol{\xi}) := ((x_i, y_i) : i \in [t], b_i = 0)$. Now, we can define a learning algorithm $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$ under bandit feedback as follows,

$$\mathbf{A}_1 : (\mathcal{X} \times \Sigma)^* \times \mathcal{X} \to \{D\} \subseteq \Pi(\mathcal{Y}) \text{ and}$$
$$\mathbf{A}_2 : (\mathcal{X} \times \Sigma)^* \times \mathcal{X} \to \mathcal{Y},\ (\boldsymbol{\xi}, x) \mapsto \mathcal{A}(\hat{\boldsymbol{z}}(\boldsymbol{\xi}))(x),$$

where $\mathbf{A}_1$ is a randomized algorithm whose output distribution is always $D \in \Pi(\mathcal{Y})$ and $\mathbf{A}_2$ is a deterministic algorithm whose output is the label $\mathcal{A}(\hat{\boldsymbol{z}}(\boldsymbol{\xi}))(x)$ for input sequence $\boldsymbol{\xi} \in (\mathcal{X} \times \Sigma)^*$ and instance $x \in \mathcal{X}$. For any $\boldsymbol{s} \in \mathcal{Z}^*$, let $\hat{h}^{\boldsymbol{s}}_{\mathbf{A}}$ denote the classifier defined by $\mathbf{A}$ under the example sequence $\boldsymbol{s}$ with bandit feedback, i.e., $\hat{h}^{\boldsymbol{s}}_{\mathbf{A}}(x) := \mathbf{A}_2(\mathbf{B}_{\mathbf{A}}(\boldsymbol{s}), x) = \mathcal{A}(\hat{\boldsymbol{z}}(\mathbf{B}_{\mathbf{A}}(\boldsymbol{s})))(x)$ for any $x \in \mathcal{X}$.

Sampling $S \sim P^{\mathbb{N}}$, by Lemma 42, we have $\hat{\boldsymbol{z}}(\mathbf{B}_{\mathbf{A}}(S)) \sim (P_D)^{\mathbb{N}}$, where $P_D \in \mathrm{RE}(\mathcal{C})$ is a distribution over $\mathcal{Z}$ such that $P_D(E) = \mathbb{P}_{((X,Y),Y')\sim P\times D}((X, Y) \in E \mid Y = Y')$ for any measurable subset E of $\mathcal{Z}$. Define $\boldsymbol{Z}_n := \hat{\boldsymbol{z}}(\mathbf{B}_{\mathbf{A}}(S_{\leqslant n}))$ and $K_n := |\boldsymbol{Z}_n| \in \mathbb{N} \cup \{0\}$ for any $n \in \mathbb{N}$. Then, we have $\lim_{n\to\infty} K_n = \infty$ almost surely since $|\hat{\boldsymbol{z}}(\mathbf{B}_{\mathbf{A}}(S))| = \infty$. Conditional on K_n, we have $\boldsymbol{Z}_n \sim (P_D)^{K_n}$. Define $r(n) := \mathbb{E}_{T\sim(P_D)^n}[\mathrm{er}_{P_D}(\mathcal{A}(T))]$ for any $n \in \mathbb{N}$. Since $P_D \in \mathrm{RE}(\mathcal{C})$, the universal learnability of $\mathcal{A}$ implies that $\lim_{n\to\infty} r(n) = 0$. By the definition of $\mathbf{A}$, we have

$$\mathbb{E}[\mathrm{er}_{P_D}(\hat{h}^{S_{\leqslant n}}_{\mathbf{A}}) \mid K_n] = \mathbb{E}[\mathrm{er}_{P_D}(\mathcal{A}(\boldsymbol{Z}_n) \mid K_n] = r(K_n).$$

Since $K_n \to \infty$ almost surely, we have $r(K_n) \to 0$ almost surely and by bounded convergence,

$$\lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[\mathrm{er}_{P_D}(\hat{h}^{S_n}_{\mathbf{A}})] = \lim_{n\to\infty} \mathbb{E}[r(K_n)] = \mathbb{E}\left[\lim_{n\to\infty} r(K_n)\right] = 0.$$

Sample $(X, Y) \sim P$ and $Y' \sim D$ independently. Define $\mathcal{Z}_y := \mathcal{X} \times \{y\}$ and $p_{P,D} := \mathbb{P}(Y = Y') = \sum_{y\in\mathcal{Y}} P(\mathcal{Z}_y)D(y) > 0$. For any $h : \mathcal{X} \to \mathcal{Y}$ and $y \in \mathcal{Y}$, we define $\mathrm{ER}(h) := \{(x, y) \in \mathcal{Z} : h(x) \neq y\}$ and $\mathrm{ER}_y(h) := \{(x, y) : x \in \mathcal{X}, h(x) \neq y\} = \mathrm{ER}(h) \cap \mathcal{Z}_y$. Then, we have $\mathrm{ER}(h) = \cup_{y\in\mathcal{Y}}\mathrm{ER}_y(h)$, $\mathrm{ER}_y(h) \cap \mathrm{ER}_{y'}(h) = \varnothing$ for any $\mathcal{Y} \ni y \neq y' \in \mathcal{Y}$, and

$$\mathrm{er}_{P_D}(h) = \frac{\mathbb{P}(h(X) \neq Y, Y = Y')}{\mathbb{P}(Y = Y')} = \frac{\sum_{y\in\mathcal{Y}} P(\mathrm{ER}_y(h))D(y)}{p_{P,D}}.$$

It follows that $P(\mathrm{ER}_y(h)) \leqslant \frac{p_{P,D}\mathrm{er}_{P_D}(h)}{D(y)}$ for all $y \in \mathcal{Y}$, which implies that

$$\mathbb{E}_{S_n\sim P^n}[P(\mathrm{ER}_y(\hat{h}^{S_n}_{\mathbf{A}})] \leqslant \frac{p_{P,D}}{D(y)}\mathbb{E}_{S_n\sim P^n}[\mathrm{er}_{P_D}(\hat{h}^{S_n}_{\mathbf{A}})] \text{ and thus}$$
$$\lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[P(\mathrm{ER}_y(\hat{h}^{S_n}_{\mathbf{A}})] = 0.$$

Since $\mathbb{E}_{S_n\sim P^n}[P(\mathrm{ER}_y(\hat{h}_{\mathbf{A}}^{S_n})] \leqslant \mathbb{E}_{S_n\sim P^n}[P(\mathcal{Z}_y)] = P(\mathcal{Z}_y)$ for any $n \in \mathbb{N}$ and $\sum_{y\in\mathcal{Y}} P(\mathcal{Z}_y) = P(\cup_{y\in\mathcal{Y}}\mathcal{Z}_y) = 1$, by the dominated convergence theorem, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[\mathrm{er}_P(\hat{h}_{\mathbf{A}}^{S_n})] &= \lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[P(\cup_{y\in\mathcal{Y}}\mathrm{ER}_y(\hat{h}_{\mathbf{A}}^{S_n}))] \\
&= \lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[\textstyle\sum_{y\in\mathcal{Y}} P(\mathrm{ER}_y(\hat{h}_{\mathbf{A}}^{S_n})] \\
&= \lim_{n\to\infty} \textstyle\sum_{y\in\mathcal{Y}} \mathbb{E}_{S_n\sim P^n}[P(\mathrm{ER}_y(\hat{h}_{\mathbf{A}}^{S_n})] \\
&= \textstyle\sum_{y\in\mathcal{Y}} \lim_{n\to\infty} \mathbb{E}_{S_n\sim P^n}[P(\mathrm{ER}_y(\hat{h}_{\mathbf{A}}^{S_n})] = 0.
\end{aligned}$$

Thus, $\mathcal{Q}$ is universal learnable under bandit feedback. ■