Complexity of Teaching by a Restricted Number of Examples

Hayato Kobayashi and Ayumi Shinohara
Graduate School of Information Sciences, Tohoku University, Japan
`kobayashi@shino.ecei.tohoku.ac.jp` and `ayumi@ecei.tohoku.ac.jp`

Abstract

Teaching is inextricably linked to learning, and there are many studies on the complexity of teaching as well as learning in computational learning theory. In this paper, we study the complexity of teaching in the situation that the number of examples is restricted, especially less than its teaching dimension. We formulate a model of teaching by a restricted number of examples, where the complexity is measured by the maximum error to a target concept. We call a concept class is *optimally incrementally teachable* if the teacher can optimally teach it to the learner whenever teaching is terminated.

We study the complexity of the three concept classes of monotone monomials, monomials without the empty concept, and monomials in our model. We show that the boundary of optimally incremental teachability is different from that of polynomial teachability in the classical model. We also show that inconsistent examples help to reduce the maximum error in our model.

1 Introduction

Computational learning theory aims to mathematically formulate a model of *learning*, which is an intellectual behavior of humans, and finally bring out its nature. In the domain, there are many studies using models such as inductive inference [Gol67], PAC learning [Val84], and query learning [Ang88]. On the other hand, there are also many studies on *teaching* [SM91, GK95, JT92, Heg95, Mat97, LSW07, Han07, Bal08, ZLHZ08], which is inextricably linked to learning. In this paper, we formulate a new model of teaching and discuss the complexity of teaching in the model.

We assume the situation that the number of examples is restricted to some integer, especially less than its teaching dimension. This situation brings teaching theory closer to the real world, since time is limited in general when we teach something to someone. For example, a teacher in a school should preferably make the students understand his/her lesson within course hours. Self-introduction in a job interview and presentation of research results are good instances of the situation, which many people may experience. In addition, interactive learning of robots is another instance of the situation, since the number of trials is often restricted due to the issue of cost.

We formulate a model of teaching in the situation that the number of examples is restricted. In our model, we measure the teaching complexity of a concept class by the minimum *teaching error*. Teaching error is informally the worst case error between a target concept and any concept consistent with a set S of examples. If the teaching error is zero, then the set S specifies the target concept uniquely, and S is called a *teaching set*. The size of the minimum teaching set was of interest in the standard model of teaching, and it is called the *teaching dimension* [GK95]. In this paper, we are interested in a situation that only a smaller number of examples are allowed to use than the teaching dimension of the target concept. The goal of the teacher is to minimize the misunderstanding of learners to the target concept, which cannot be zero, by selecting a good set S' of examples. Some questions arise. Is an optimal teaching set in this situation always a subset of the standard teaching set, i.e. $S' \subseteq S$? Is it possible to reduce the teaching error by allowing the teacher to use inconsistent examples? We prove some theorems which answer these questions. Theorem 3 in Section 3 answers to the first question negatively. We show a simple and concrete concept class that the optimal teaching set in restricted numbers of examples is not a subset of the standard teaching set. It implies that a good teaching strategy within a restricted number of examples is different from the standard one. Theorem 4 answers to the second question affirmatively. We show another simple and concrete concept class for which the optimal teaching set is inconsistent with the target concept. It supports our intuitions that *it is sometimes necessary to lie*.

Moreover, we define that a concept is *optimally incrementally teachable*, if there exists a sequence of examples such that any prefix of the sequence is always an optimal teaching set at that moment. When a concept is optimally incrementally teachable, it should be easy to teach in our model in the sense that we do not need to change its teaching algorithm depending on the restriction of the number of examples. We address the three concept classes of monotone monomials, and monomials without the empty concept, monomials and bring out their properties in our model.

Our model is inspired by the study of Balbach and Zeugmann [BZ06]. They formalized a model for randomized learners and measured the complexity of a concept class by its expected teaching time. In the study, interestingly, they showed that inconsistent teachers can be more powerful than consistent teachers using self-made concept classes with only positive examples. It supports the same intuitions that it is sometimes necessary to lie.

Our model is not categorized in exact learning differently from most other studies and relatively closer to PAC learning. However, our model has not the notion of probability as in PAC learning. In this paper, we adopt the style assuming the worst case learner in the same way as the classical model [SM91, GK95].

2 Preliminaries

We denote the size of a set S by either $\#S$ or $|S|$. Let **N** be a set of natural numbers. We define $[i,j] := \{\, i, i+1, \cdots, j\,\}$ for any integers $i, j \in \mathbf{N}$ with $i \leq j$.

Let Σ be an alphabet. For any string $u \in \Sigma^n$, we denote a character at a position $i \in [1,n]$ by $u[i]$, and we define $u[i:j] := u[i]u[i{+}1]\cdots u[j{-}1]$ for any integers $i, j \in [1,n]$ with $i \leq j$. For any two strings $u_1, u_2 \in \Sigma^n$ of the same length n, we denote the Hamming distance between u_1 and u_2 by $H(u_1, u_2)$.

We now consider the case $\Sigma = \{0,1\}$. For any binary string $u \in \Sigma^n$, we denote the set of binary strings such that its Hamming distance to u is 1 by

$$\mathcal{N}_1(u) := \{u' \in \Sigma^n \mid H(u,u') = 1\}.$$

2.1 Concept Class

Let X be a set of *instances*, or an *input space*. Let $\mathcal{X} := X \times \{0,1\}$, and we call a pair $(x,b) \in \mathcal{X}$ an *example*. We call $\mathcal{C} \subseteq 2^X$ a *concept class* and $c \in \mathcal{C}$ a *concept*. A concept c is *consistent* with an example $(x,b) \in \mathcal{X}$ iff $(x,b) \in \mathcal{X}(c)$, where

$$\mathcal{X}(c) := \{(x,b) \in \mathcal{X} \mid x \in c \Leftrightarrow b = 1\}.$$

We denote the set of concepts in a concept class $\mathcal{C}$ that are consistent with a set S of examples by

$$\mathit{CONS}(S,C) := \{c \in \mathcal{C} \mid S \subseteq \mathcal{X}(c)\}.$$

For any two concepts $c_1, c_2 \in \mathcal{C}$, we denote the *symmetric difference* between c_1 and c_2 by

$$c_1 \Delta c_2 := (c_1 - c_2) \cup (c_2 - c_1).$$

For any two concepts $c_1, c_2 \in \mathcal{C}$, the *error* between c_1 and c_2 with respect to a distribution D is calculated by $d_D(c_1,c_2) := \sum_{x \in c_1 \Delta c_2} \Pr_D(x)$. In this paper, we assume that an input space X is finite, and a distribution D is uniform, so that we omit D and calculate the error by the following simple formula

$$d(c_1,c_2) := \frac{|c_1 \Delta c_2|}{|X|}.$$

2.2 Teaching Dimension

A set S of examples is called a *teaching set* of a concept c with respect to a concept class $\mathcal{C}$ if c is the only concept in $\mathcal{C}$ that is consistent with S, that is $\mathit{CONS}(S,\mathcal{C}) = \{c\}$. We denote the set of all teaching sets by

$$\mathit{TS}(c,\mathcal{C}) := \{S \subseteq \mathcal{X} \mid \mathit{CONS}(S,\mathcal{C}) = \{c\}\}.$$

The *teaching dimension* of c with respect to $\mathcal{C}$ is defined by the minimum size of its teaching set:

$$\mathit{TD}(c,\mathcal{C}) := \min\{|S| \mid S \in \mathit{TS}(c,\mathcal{C})\}.$$

The teaching dimension of $\mathcal{C}$ is defined by the maximum teaching dimension over all concepts:

$$\mathit{TD}(\mathcal{C}) := \max_{c \in \mathcal{C}} \mathit{TD}(c,\mathcal{C}).$$

We denote the set of all minimum teaching sets of c with respect to $\mathcal{C}$ by

$$\mathit{MinTS}(c,\mathcal{C}) := \{S \in \mathit{TS}(c,S) \mid |S| = \mathit{TD}(c,\mathcal{C})\}.$$

3 Teaching Complexity by a Restricted Number of Examples

In our model, a learner often fails to identify a target concept exactly, since the number of examples that a teacher can give to the learner is restricted. We measure the teaching complexity of a concept class by the worst case error of the target concept. In this case, we assume the worst case learner in the same way as the classical teaching model. We define the worst case error of the target concept as follows.

Definition 1 (Teaching Error)**.** *Let $\mathcal{C}$ be a concept class and $c \in \mathcal{C}$ be a target concept. For any set S of examples, we define* the teaching error of c with respect to $\mathcal{C}$ by S *as*

$$\mathit{TE}(c,\mathcal{C},S) := \begin{cases} \max\limits_{c' \in \mathit{CONS}(S,\mathcal{C})} d(c,c') & (\mathit{CONS}(S,\mathcal{C}) \neq \emptyset), \\ 1 & (otherwise). \end{cases}$$

We define the teaching complexity of a concept class in the situation that the number of examples is restricted as follows.

Definition 2 (Optimal Teaching Error)**.** *Let $\mathcal{C}$ be a concept class and $c \in \mathcal{C}$ be a target concept. We define* the optimal teaching error of c with respect to $\mathcal{C}$ by at most k examples *as*

$$\mathit{OptTE}_k(c,\mathcal{C}) := \min_{S \subseteq \mathcal{X} : |S| \leq k} \mathit{TE}(c,\mathcal{C},S).$$

We also define the optimal teaching error of $\mathcal{C}$ by at most k examples *as*

$$\mathit{OptTE}_k(\mathcal{C}) := \max_{c \in \mathcal{C}} \mathit{OptTE}_k(c,\mathcal{C}).$$

We define that a set of at most k examples is a *k-optimal teaching set* if the set achieves the optimal teaching error. We denote the set of k-optimal teaching sets by

$$\mathit{OptTS}_k(c,\mathcal{C}) := \left\{ S \subseteq \mathcal{X} \;\middle|\; \begin{array}{l} |S| \leq k, \\ \mathit{TE}(c,\mathcal{C},S) = \mathit{OptTE}_k(c,\mathcal{C}) \end{array} \right\}.$$

An optimal teaching error ranges between 0 and 1, and the smaller the better. When $k \geq \mathit{TD}(c,\mathcal{C})$, the optimal teaching error is always zero, i.e., $\mathit{OptTE}_k(c,\mathcal{C}) = 0$, since we can uniquely specify the target concept by any teaching set S in $\mathit{MinTS}(c,\mathcal{C})$. Therefore, this paper focuses on the case of $k < \mathit{TD}(c,\mathcal{C})$.

We now prove the next two important theorems.

Table 1: Concept class used in the proof for Theorem 3.

h	x_1	x_2	x_3	$d(c_0, h)$
c_0	1	1	1	0/3
c_1	1	0	1	1/3
c_2	1	0	0	2/3
c_3	0	1	1	1/3
c_4	0	1	0	2/3

Theorem 3. *There exists a concept class $\mathcal{C}$, a target concept $c \in \mathcal{C}$, and a positive integer k such that*

$$\forall S_1 \in \mathit{OptTS}_k(c, \mathcal{C}),\ \forall S_2 \in \mathit{MinTS}(c, \mathcal{C}), \quad S_1 \nsubseteq S_2.$$

Proof: Let us consider a concept class $\mathcal{C} = \{c_0, \cdots, c_4\}$ over an input space $X = \{x_1, x_2, x_3\}$ shown in Table 1 and a target concept $c_0 \in \mathcal{C}$, and let us take $k = 1$.

The set $S_1 := \{(x_3, 1)\}$ is the only 1-optimal teaching set of c_0 with respect to $\mathcal{C}$. That is

$$\mathit{OptTS}_1(c_0, \mathcal{C}) = \{S_1\}.$$

Its optimal teaching error is $\mathit{OptTE}_1(c_0, \mathcal{C}) = 1/3$. For any other set S of examples, S is not optimal since either $c_2 \in \mathit{CONS}(S, \mathcal{C})$ or $c_4 \in \mathit{CONS}(S, \mathcal{C})$, i.e., $\mathit{TE}(c_0, \mathcal{C}, S) \geq 2/3$. On the other hand, the set $S_2 := \{(x_1, 1), (x_2, 1)\}$ is the only minimum teaching set. That is

$$\mathit{MinTS}(c_0, \mathcal{C}) = \{S_2\}.$$

However,

$$S_1 \nsubseteq S_2.$$

■

Theorem 4. *There exists a concept class $\mathcal{C}$, a target concept $c \in \mathcal{C}$, and a positive integer k such that*

$$\forall S \in \mathit{OptTS}_k(c, \mathcal{C}), \quad c \notin \mathit{CONS}(S, \mathcal{C}).$$

Proof: Let us consider a concept class $\mathcal{C} = \{c_0, \cdots, c_5\}$ over an input space $X = \{x_1, \cdots, x_5\}$ shown in Table 2 and a target concept $c_0 \in \mathcal{C}$, and let us take $k = 1$.

The set $S_1 := \{(x_1, 0)\}$ is the only 1-optimal teaching set. That is

$$\mathit{OptTS}_1(c_0, \mathcal{C}) = \{S_1\}.$$

Its optimal teaching error is $\mathit{OptTE}_k(c_0, \mathcal{C}) = 1/5$. For any other set S of examples, S is not optimal, since $c_i \in \mathit{CONS}(S, \mathcal{C})$ for some $i \in [2, 5]$, i.e., $\mathit{TE}(c_0, \mathcal{C}, S) \geq 3/5$. However, c_0 is inconsistent with S_1, that is

$$c_0 \notin \mathit{CONS}(S_1, \mathcal{C}).$$

■

Theorem 3 and Theorem 4 indicate that a teaching strategy in our model can be much different from that in the classical model. In other words, we need an appropriate teaching algorithm to obtain the maximum teaching effect in the situation that the number of examples is restricted. Although we use simple self-made concept classes in the proofs of Theorem 3 and Theorem 4 for illustrative purposes, the natural concept class of monomials also has the same properties as them, as shown in Theorem 36.

If for a concept class, there exists a teaching algorithm that achieves the optimal teaching error whenever teaching is terminated, we can regard such a concept class as somewhat easy to teach in our model. We formalize the idea as follows.

Table 2: Concept class used in the proof for Theorem 4.

h	x_1	x_2	x_3	x_4	x_5	$d(c_0, h)$
c_0	1	1	1	1	1	0/5
c_1	0	1	1	1	1	1/5
c_2	1	1	0	0	0	3/5
c_3	1	0	1	0	0	3/5
c_4	1	0	0	1	0	3/5
c_5	1	0	0	0	1	3/5

Definition 5 (Optimally Incremental Teachability). *Let $\mathcal{C}$ be a concept class and $c \in \mathcal{C}$ be a target concept. We say that c* is optimally incremental teachable with respect to $\mathcal{C}$ *if there exists a list $L := \langle z_1, z_2, \cdots \rangle$ of examples such that*

$$\forall k \in [1, \mathit{TD}(c, \mathcal{C})], \{z_1, \cdots, z_k\} \in \mathit{OptTS}_k(c, \mathcal{C}).$$

We also say that $\mathcal{C}$ is optimally incrementally teachable *if every $c \in \mathcal{C}$ is optimally incrementally teachable with respect to $\mathcal{C}$.*

4 Complexity of Teaching Monomials

4.1 Preliminaries

We deal with Boolean concepts over an input space $X_n = \{0, 1\}^n$, expressed by Boolean expressions. A concept expressed by a Boolean expression f is the set of inputs in X_n which satisfies f. A *monomial* on n variables $v_1, \cdots, v_n$ is a Boolean expression represented by a conjunction of some literals, v_i's and $\overline{v_i}$'s. A concept represented by an unsatisfiable expression, such as $v_i \wedge \overline{v_i}$ is called the *empty concept* and denoted by $c_e := \emptyset$.

We denote by $\mathcal{M}_n$ the class of all monomial concepts. Let $\mathcal{M}'_n := \mathcal{M}_n - \{c_e\}$. Any concept $c \in \mathcal{M}'_n$ is uniquely represented by a string $r \in \{0, 1, *\}^n$ defined as follows. $r[i]$ is 1 if there exists v_i in a corresponding Boolean expression, 0 if there exists $\overline{v_i}$, and $*$ otherwise. For instance, for the monomial $v_1 \wedge \overline{v_2}$ on three variables, the corresponding concept c is $\{100, 101\}$, and its representation is $10*$. For the sake of simplicity, we identify a concept in $\mathcal{M}'_n$ with its representation:

$$c = \{100, 101\} = 10*.$$

For any concept $c \in \mathcal{M}'_n$, we denote the number of distinct variables in a corresponding Boolean expression by

$$var(c) := \#\{i \mid c[i] \neq *\}.$$

A monomial consisting of only positive literals is called a *monotone monomial*. We denote the concept class of monotone monomials by $\mathcal{M}^+_n$. Obviously $\mathcal{M}^+_n \subseteq \mathcal{M}'_n$.

For any two concepts $c_1, c_2 \in \mathcal{M}'_n$, we say that c_1 *and* c_2 *have a strong difference at position* i if either $c_1[i] = 1$ and $c_2[i] = 0$, or $c_1[i] = 0$ and $c_2[i] = 1$. We denote the number of strong differences between c_1 and c_2 by

$$s(c_1, c_2) := \#\{i \mid c_1[i] \neq *, c_2[i] \neq *, c_1[i] \neq c_2[i]\}.$$

We say that c_1 *and* c_2 *have a weak difference at position* i if either $c_1[i] = *$ and $c_2[i] \in \{0, 1\}$, or $c_2[i] = *$ and $c_1[i] \in \{0, 1\}$. Two weak differences at position i and j are called *the same kind* if either $c_1[i] = c_1[j] = *$ or $c_2[i] = c_2[j] = *$, and called *different kind* otherwise. We denote the number of weak differences of the same kind between c_1 and c_2 by

$$w(c_1, c_2) := \#\{i \mid c_1[i] \neq *, c_2[i] = *\}.$$

The total number of weak differences between c_1 and c_2 is $w(c_1, c_2) + w(c_2, c_1)$. We say that c_1 *and* c_2 *have an arbitrary match at position* i if $c_1[i] = c_2[i] = *$. We denote the number of arbitrary matches by

$$a(c_1, c_2) := \#\{i \mid c_1[i] = c_2[i] = *\}.$$

4.2 Monotone Monomials

Shinohara and Miyano [SM91], and Goldman and Kearns [GK95] independently studied the concept class of monotone monomials and proved that its teaching dimension is n. Goldman and Kearns moreover analyzed the teaching dimension for every concept in the class $\mathcal{M}_n^+$ and proved the following theorem.

Theorem 6 ([GK95])**.** *For any concept $c \in \mathcal{M}_n^+$, the teaching dimension of c with regard to $\mathcal{M}_n^+$ is calculated as*

$$TD(c, \mathcal{M}_n^+) = \min\{var(c) + 1, n\}.$$

The set S_ℓ of examples defined by the following formula is one of the minimum teaching sets of the concept $1^\ell *^{n-\ell}$ with regard to $\mathcal{M}_n^+$.

$$S_\ell := \begin{cases} \{(u, 0) \mid u \in \mathcal{N}_1(1^\ell)\} & (\ell = n), \\ \{(u1^{n-\ell}, 0) \mid u \in \mathcal{N}_1(1^\ell)\} \cup \{(1^\ell 0^{n-\ell}, 1)\} & (\ell < n). \end{cases}$$

From the view point of our study, the above theorem describes the complexity of teaching without restriction of the number of examples to use. On the other hand, this paper focuses on the complexity of teaching by a restricted number of examples.

The next lemma shows the formula calculating the error between any two concepts in $\mathcal{M}_n^+$. Note that we do not consider the exceptions for $n \leq 2$ in our proofs, assuming that n is relatively large.

Lemma 7. *For any two concepts $c_1, c_2 \in \mathcal{M}_n^+$,*

$$d(c_1, c_2) = \frac{(2^{\tilde{w}_1} + 2^{\tilde{w}_2} - 2)2^{\tilde{a}}}{2^n},$$

where $\tilde{w}_1 = w(c_1, c_2)$, $\tilde{w}_2 = w(c_2, c_1)$, and $\tilde{a} = a(c_1, c_2)$.

Proof: By the definition, $|c_1 \Delta c_2|$ is the number of instances $x \in X_n$ such that either $x \in c_1$ and $x \notin c_2$, or $x \notin c_1$ and $x \in c_2$.

We first count the number of instances x such that $x \in c_1$ and $x \notin c_2$. The number of instances $x \in c_1$ is $2^{\tilde{w}_2 + \tilde{a}}$, where $\tilde{w}_2 + \tilde{a}$ means the number of $*$ in c_1, since for any integer $i \in [1, n]$,

$$c_1[i] \neq * \quad \Rightarrow \quad x[i] = c_1[i],$$

in order to satisfy the Boolean expression corresponding to c_1. In those instances, however, there are $2^{\tilde{a}}$ instances such that $x \in c_2$, since the Boolean expression corresponding to c_2 is satisfied if for any $i \in [1, n]$,

$$c_2[i] \neq * \quad \Rightarrow \quad x[i] = c_2[i].$$

Therefore, the number of instances x such that both $x \in c_1$ and $x \notin c_2$ is $(2^{\tilde{w}_2} - 1)2^{\tilde{a}}$.

In the same way, we can verify that the number of instances x such that both $x \notin c_1$ and $x \in c_2$ is $(2^{\tilde{w}_1} - 1)2^{\tilde{a}}$.

The formula for $d(c_1, c_2)$ is obtained by adding $(2^{\tilde{w}_2} - 1)2^{\tilde{a}}$ and $(2^{\tilde{w}_1} - 1)2^{\tilde{a}}$, and dividing it by $|X_n| = 2^n$. ■

We prove the following five lemmas for the proof of Theorem 13 which exactly shows the optimal teaching error of $\mathcal{M}_n^+$.

Lemma 8. *For any instance $x \in \mathcal{N}_1(1^n)$ and any integer $i \in [1, n]$,*

$$x[i] = 0 \quad \Rightarrow \quad \forall c \in \mathit{CONS}(\{(x, 0)\}, \mathcal{M}_n^+), c[i] = 1$$

Proof: Suppose that

$$\exists c \in \mathit{CONS}(\{(x, 0)\}, \mathcal{M}_n^+), \quad c[i] = * \neq 1.$$

Since $x \in \mathcal{N}_1(1^n)$ and $x[i] = 0$, the instance x satisfies the Boolean expression corresponding to c. In other words, $c \notin \mathit{CONS}(\{(x, 0)\}, \mathcal{M}_n^+)$. This is a contradiction. ■

Lemma 9. *Let S be a set of examples with $\mathit{CONS}(S, \mathcal{M}_n^+) \neq \emptyset$. For any integer $k \in [1, n]$,*

$$|S^-| < k \quad \Rightarrow \quad \exists c \in \mathit{CONS}(S, \mathcal{M}_n^+), var(c) < k,$$

where $S^- := \{(x, b) \in S \mid b = 0\}$.

Proof: We prove the contraposition

$$\left[\forall c \in \mathit{CONS}(S, \mathcal{M}_n^+),\ var(c) \geq k\right] \quad \Rightarrow \quad |S^-| \geq k.$$

Since $\mathit{CONS}(S, \mathcal{M}_n^+) \neq \emptyset$, there exists at least one concept c satisfying the assumption of the contraposition. We define such a concept as $c := 1^\ell *^{n-\ell}$ without loss of generality, where $\ell := var(c)$.

Let $c_i := 1^{i-1} * 1^{k-i} *^{n-k}$ for $i \in [1, k]$. By the definition, $var(c_i) = k - 1$. From the assumption of the contraposition, $c_i \notin \mathit{CONS}(S, \mathcal{M}_n^+)$. That is, $\exists z \in S, z \notin \mathcal{X}(c_i)$. On the other hand, $\forall z \in S, z \in \mathcal{X}(c)$, since $c \in \mathit{CONS}(S, \mathcal{M}_n^+)$. An example z_i such that $z_i \notin \mathcal{X}(c_i)$ and $z_i \in \mathcal{X}(c)$ must be negative, since $c \subseteq c_i$. Then, z_i must be of the form

$$z_i = (1^{i-1} 0 1^{k-i} u, 0),$$

where $u \in \{0, 1\}^{n-k}$. Therefore,

$$|S^-| \geq \#\{z_i \mid i \in [1, k]\} = k.$$

■

Lemma 10. *Let $c := 1^\ell *^{n-\ell} \in \mathcal{M}_n^+$. For any integer $k \in [1, \mathit{TD}(c, \mathcal{M}_n^+) - 1]$ and the set $S := \{(u1^{n-k}, 0) \mid u \in \mathcal{N}_1(1^k)\}$,*

$$\mathit{TE}(c, \mathcal{M}_n^+, S) \leq \begin{cases} \dfrac{2^{n-k} - 2^{n-\ell}}{2^n} & (k < \ell), \\ \dfrac{2^{n-k} - 1}{2^n} & (k = \ell). \end{cases}$$

Proof: By Lemma 8,

$$\mathit{CONS}(S, \mathcal{M}_n^+) = \{1^k u \mid u \in \{1, *\}^{n-k}\}.$$

For any $c' \in \mathit{CONS}(S, \mathcal{M}_n^+)$,

$$\begin{aligned} 0 \leq a(c, c') &\leq n - \ell, \\ 0 \leq w(c, c') &\leq \ell - k, \text{ and} \\ a(c, c') + w(c', c) &= n - \ell. \end{aligned}$$

Therefore,

$$\begin{aligned} TE(c,\mathcal{M}_n^+,S) &\le \max_{\substack{0\le\tilde{a}\le n-\ell\\0\le\tilde{w}\le\ell-k}} \frac{(2^{\tilde{w}}+2^{n-\ell-\tilde{a}}-2)2^{\tilde{a}}}{2^n} \\ &= \begin{cases} \dfrac{2^{n-k}-2^{n-\ell}}{2^n} & (k<\ell) \\ \dfrac{2^{n-k}-1}{2^n} & (k=\ell) \end{cases} \end{aligned}$$

where $\tilde{a} := a(c,c')$ and $\tilde{w} := w(c,c')$. ∎

Lemma 11. *For any concept $c \in \mathcal{M}_n^+$ and any integer $k \in [1, TD(c,\mathcal{M}_n^+)-1]$,*

$$\forall S \in OptTS_k(c,\mathcal{M}_n^+),\ \forall(x,b)\in S,\quad b=0.$$

Proof: Suppose that

$$\exists S \in OptTS_k(c,\mathcal{M}_n^+),\ \exists(x,b)\in S,\quad b=1.$$

Let $S^- := \{(x,b)\in S \mid b=0\}$. Obviously, $|S^-|<k$. Since $S \in OptTS_k(c,\mathcal{M}_n^+)$, we have $CONS(S,\mathcal{M}_n^+)\neq\emptyset$. By Lemma 9,

$$\exists c' \in CONS(S,\mathcal{M}_n^+),\quad var(c')<k.$$

Let $\ell' := var(c')$. For the concept c',

$$\begin{aligned} a(c,c')+w(c,c') &= n-\ell', \\ a(c,c')+w(c',c) &= n-\ell, \text{ and} \\ a(c,c') &\le n-\ell. \end{aligned}$$

Therefore, we obtain a lower bound of $OptTE_k(c,\mathcal{M}_n^+)$ as

$$\begin{aligned} OptTE_k(c,\mathcal{M}_n^+) &\ge d(c,c') \\ &\ge \min_{\tilde{a}\le n-\ell} \frac{(2^{n-\ell'-\tilde{a}}+2^{n-\ell-\tilde{a}}-2)2^{\tilde{a}}}{2^n} \\ &= \frac{2^{n-\ell'}-2^{n-\ell}}{2^n}, \end{aligned}$$

where $\tilde{a} := a(c,c')$.

When $k<\ell$. By Lemma 10, we obtain an upper bound of $OptTE_k(c,\mathcal{M}_n^+)$ as

$$OptTE_k(c,\mathcal{M}_n^+) \le \frac{2^{n-k}-2^{n-\ell}}{2^n} < \frac{2^{n-\ell'}-2^{n-\ell}}{2^n},$$

since $\ell'<k$. This is a contradiction.

When $\ell=k$. By Lemma 10, we obtain an upper bound of $OptTE_k(c,\mathcal{M}_n^+)$ as

$$OptTE_k(c,\mathcal{M}_n^+) \le \frac{2^{n-k}-1}{2^n} < \frac{2^{n-\ell}}{2^n} \le \frac{2^{n-\ell'}-2^{n-\ell}}{2^n},$$

since $\ell'<k=\ell$. This is a contradiction. ∎

Lemma 12. *For any concept $c \in \mathcal{M}_n^+$ and any integer $k \in [1, TD(c,\mathcal{M}_n^+)-1]$,*

$$OptTE_k(c,\mathcal{M}_n^+) = \begin{cases} \dfrac{2^{n-k}-2^{n-\ell}}{2^n} & (k<\ell), \\ \dfrac{2^{n-k}-1}{2^n} & (k=\ell), \end{cases}$$

where $\ell := var(c)$.

Proof: Let $c = 1^\ell *^{n-\ell}$ without loss of generality. The formula in Lemma 10 gives an upper bound of $OptTE_k(c,\mathcal{M}_n^+)$. We show that a lower bound of $OptTE_k(c,\mathcal{M}_n^+)$ is equal to the upper bound.

In the case $\ell\neq k$, for any $S\in OptTS_k(c,\mathcal{M}_n^+)$, we have $CONS(S,\mathcal{M}_n^+)\neq\emptyset$ and $|S|<k+1$. By Lemma 9,

$$\exists c' \in CONS(S,\mathcal{M}_n^+),\quad var(c')<k+1.$$

Let $\ell' := var(c')$. Obviously $\ell'\le k<k+1$. By using the same argument in Lemma 11, we obtain a lower bound of $OptTE_k(c,\mathcal{M}_n^+)$ as

$$OptTE_k(c,\mathcal{M}_n^+) \ge \frac{2^{n-\ell'}-2^{n-\ell}}{2^n} \ge \frac{2^{n-k}-2^{n-\ell}}{2^n}.$$

In the case $\ell=k$, for any $S\in OptTS_k(c,\mathcal{M}_n^+)$, $(1^n,0)\notin S$ since $CONS(S,\mathcal{M}_n^+)\neq\emptyset$. From Lemma 11 and $(1^n,0)\notin S$,

$$c_1 := 1^n \in CONS(S,\mathcal{M}_n^+).$$

For the concept c_1,

$$\begin{aligned} a(c,c_1) &= 0, \\ w(c,c_1) &= 0, \quad \text{and} \\ w(c_1,c) &= n-\ell. \end{aligned}$$

Then we obtain an lower bound of $OptTE_k(c,\mathcal{M}_n^+)$ as

$$OptTE_k(c,\mathcal{M}_n^+) \ge d(c,c_1) = \frac{2^{n-\ell}-1}{2^n} = \frac{2^{n-k}-1}{2^n}.$$

∎

The next theorem shows the optimal teaching error of $\mathcal{M}_n^+$.

Theorem 13. *For any integer $k\in[1,TD(\mathcal{M}_n^+)-1]$,*

$$OptTE_k(\mathcal{M}_n^+) = \frac{2^{n-k}-1}{2^n}.$$

Proof: Let c_k be a concept with $var(c_k)=k$. By Lemma 12, for any concept $c\in\mathcal{M}_n^+$,

$$OptTE_k(c_k,\mathcal{M}_n^+) \ge OptTE_k(c,\mathcal{M}_n^+).$$

∎

The next theorem shows the optimally incremental teachability of $\mathcal{M}_n^+$.

Theorem 14. *The concept class $\mathcal{M}_n^+$ of monotone monomials is optimally incrementally teachable.*

Proof: We prove that for any concept $c\in\mathcal{M}_n^+$, c is optimally incrementally teachable. Let $c=1^\ell *^{n-\ell}$ without loss of generality, where $\ell := var(c)$.

We define a list $\langle z_1, z_2, \cdots\rangle$ of examples as

$$z_i := \begin{cases} (1^{i-1}01^{n-i},0) & (i\le\ell), \\ (1^n,1) & (i>\ell). \end{cases}$$

When $k\in[1,TD(c,\mathcal{M}_n^+)-1]$, since the set $\{z_i \mid i\in[1,k]\}$ is equal to S in Lemma 10,

$$\{z_i \mid i\in[1,k]\} \in OptTS_k(c,\mathcal{M}_n^+).$$

When $k=TD(c,\mathcal{M}_n^+)$, from Theorem 6,

$$\{z_i \mid i\in[1,k]\} \in MinTS(c,\mathcal{M}_n^+) = OptTS_k(c,\mathcal{M}_n^+).$$

Therefore, every $c\in\mathcal{M}_n^+$ is optimally incrementally teachable. ∎

4.3 Monomials Without the Empty Concept

Goldman and Kearns [GK95] proved the following theorem that the teaching dimension of $\mathcal{M}'_n$ is $n+1$. The set S_ℓ in the theorem is obtained by adding one positive example to the set S_ℓ in Theorem 6.

Theorem 15 ([GK95])**.** *For any $c \in \mathcal{M}'_n$, the teaching dimension is calculated by*

$$TD(c, \mathcal{M}'_n) = \min\{var(c) + 2, n+1\}.$$

The following set S_ℓ is one of the minimum teaching sets for a concept $c = u*^{n-\ell}$ such that $u \in \{0,1\}^\ell$.

$$S_\ell := \begin{cases} \{(u', 0) \mid u' \in \mathcal{N}_1(u)\} \cup \{(u, 1)\} & (\ell = n), \\ \{(u'1^{n-\ell}, 0) \mid u' \in \mathcal{N}_1(u)\} \\ \quad \cup\{(u0^{n-\ell}, 1)\} \cup \{(u1^{n-\ell}, 1)\} & (\ell < n). \end{cases}$$

The next lemma shows a formula calculating the error between any two concepts in $\mathcal{M}'_n$.

Lemma 16. *For any two concepts $c_1, c_2 \in \mathcal{M}'_n$,*

$$d(c_1, c_2) = \frac{(2^{\tilde{w}_1} + 2^{\tilde{w}_2} - p(\tilde{s}))2^{\tilde{a}}}{2^n},$$

where $\tilde{s} = s(c_1, c_2)$, $\tilde{w}_1 = w(c_1, c_2)$, $\tilde{w}_2 = w(c_2, c_1)$, $\tilde{a} = a(c_1, c_2)$, and

$$p(\tilde{s}) = \begin{cases} 2 & (\tilde{s} = 0), \\ 0 & (\tilde{s} > 0). \end{cases}$$

Proof: When $\tilde{s} = 0$, by the same argument in Lemma 7, the formula of $d(c_1, c_2)$ is obtained.

When $\tilde{s} > 0$, we count the number of instances $x \in X_n$ such that $x \in c_1$ and $x \notin c_2$. Using the same argument in Lemma 7, there are $2^{\tilde{w}_2}2^{\tilde{a}}$ instances such that $x \in c_1$. In this case, since c_1 and c_2 have a strong difference at the some positions, there is no instance such that $x \in c_2$. Therefore, the number of instances x such that $x \in c_1$ and $x \notin c_2$ is $2^{\tilde{w}_2}2^{\tilde{a}}$.

In the same way, the number of instances x such that $x \notin c_1$ and $x \in c_2$ is $2^{\tilde{w}_1}2^{\tilde{a}}$.

Therefore, when $\tilde{s} > 0$, the formula of $d(c_1, c_2)$ is obtained by adding $2^{\tilde{w}_2}2^{\tilde{a}}$ and $2^{\tilde{w}_1}2^{\tilde{a}}$, and dividing it by $|X_n| = 2^n$. ■

We prove the following six lemmas for the proof of Theorem 23 which shows the optimal teaching error of $\mathcal{M}'_n$.

Lemma 17. *For any set S^- of negative examples,*

$$|S^-| < k \quad \Rightarrow \quad \exists c \in CONS(S^-, \mathcal{M}'_n),\ var(c) = \lceil \log k \rceil.$$

Proof: We prove the contraposition

$$\forall c \in CONS(S^-, \mathcal{M}'_n),\ var(c) \neq \lceil \log k \rceil \quad \Rightarrow \quad |S^-| \geq k.$$

Let $c_u := u*$ such that $u \in \{0,1\}^{\lceil \log k \rceil}$. By the assumption of the contraposition, $c_u \notin CONS(S^-, \mathcal{M}_n)$. In this case, S^- must contain a negative example $(uu', 0)$ such that $u' \in \{0,1\}^{n - \lceil \log k \rceil}$. Therefore,

$$\begin{aligned} |S^-| &\geq \#\{(u1^{n-\lceil \log k \rceil}, 0) \mid u \in \{0,1\}^{\lceil \log k \rceil}\} \\ &= 2^{\lceil \log k \rceil} \geq 2^{\log k} = k. \end{aligned}$$

■

Lemma 18. *For any set S^- of negative examples and any concept c with $\ell := var(c) \geq 1$,*

$$|S^-| < 2^n - 2^{n-\ell} \Rightarrow \exists c' \in CONS(S^-, \mathcal{M}'_n), s(c, c') \geq 1.$$

Proof: We prove the contraposition

$$\forall c' \in CONS(S^-, \mathcal{M}'_n), s(c, c') = 0 \Rightarrow |S^-| \geq 2^n - 2^{n-\ell}.$$

Let $c := 1^\ell *^{n-\ell}$ without loss of generality. From the assumption of the contraposition, for any string $u \in \{0,1\}^\ell - \{1^\ell\}$ and $u' \in \{0,1\}^{n-\ell}$,

$$uu' \notin CONS(S^-, \mathcal{M}'_n).$$

Therefore,

$$\begin{aligned} |S^-| &\geq \#\{uu' | u \in \{0,1\}^\ell - \{1^\ell\}, u' \in \{0,1\}^{n-\ell}\} \\ &= (2^\ell - 1)2^{n-\ell} = 2^n - 2^{n-\ell}. \end{aligned}$$

■

Lemma 19. *For any concept $c \in \mathcal{M}'_n$ with $\ell := var(c) \geq 1$, any set S^- of negative examples, and any positive integer $k \leq 2^n - 2^{n-\ell}$.*

$$|S^-| < k \quad \Rightarrow \quad TE(c, \mathcal{M}'_n, S^-) \geq \frac{2^{n-\lceil \log k \rceil} + 2^{n-\ell}}{2^n}.$$

Proof: From Lemma 17,

$$\exists c_k \in CONS(S^-, \mathcal{M}'_n), \quad var(c_k) = \lceil \log k \rceil.$$

For two concepts c and c_k,

$$\begin{aligned} a(c, c_k) + w(c, c_k) &= n - \lceil \log k \rceil, \\ a(c, c_k) + w(c_k, c) &= n - \ell, \qquad \text{and} \\ n - \ell - \lceil \log k \rceil \leq a(c, c_k) &\leq n - \ell. \end{aligned}$$

By Lemma 18,

$$s(c, c_k) \geq 1.$$

Therefore,

$$\begin{aligned} TE(c, \mathcal{M}'_n, S^-) &\geq d(c, c_k) \\ &= \frac{(2^{n-\lceil \log k \rceil - \tilde{a}} + 2^{n-\ell-\tilde{a}})2^{\tilde{a}}}{2^n} \\ &= \frac{2^{n-\lceil \log k \rceil} + 2^{n-\ell}}{2^n}, \end{aligned}$$

where $\tilde{a} := a(c, c_k)$. ■

Lemma 20. *Let $c := u*^{n-\ell} \in \mathcal{M}'_n$ such that $u \in \{0,1\}^\ell$. For any integer $k \in [1, TD(c, \mathcal{M}'_n) - 1]$ and the set $S := \{(u1^{n-\ell}, 1)\} \cup \{(u'u[k:\ell{+}1]1^{n-\ell}, 0) \mid u' \in \mathcal{N}_1(u[1:k])\}$,*

$$TE(c, \mathcal{M}'_n, S) \leq \begin{cases} \dfrac{2^{n-k+1} - 2^{n-\ell}}{2^n} & (k < \ell + 1), \\ \dfrac{2^{n-k+1} - 1}{2^n} & (k = \ell + 1). \end{cases}$$

Proof: For any positive example $z^+ := (x, 1) \in \mathcal{X}$, we can identify $CONS(\{z^+\}, \mathcal{M}'_n)$ with $\mathcal{M}^+_n$, since

$$\forall c \in CONS(\{z^+\}, \mathcal{M}'_n), \forall i \in [1, n], c[i] \in \{x[i], *\}.$$

Therefore, we obtain the formula by using the same argument in Lemma 10. ■

Lemma 21. *For any concept $c \in \mathcal{M}'_n$ with $\ell := var(c) \geq 2$ and any integer $k \in [3, TD(c, \mathcal{M}'_n) - 1]$,*

$$\forall S \in OptTS_k(c, \mathcal{M}'_n),\ \exists (x,b) \in S, \quad b = 1.$$

Proof: Suppose that

$$\exists S \in OptTS_k(c, \mathcal{M}'_n),\ \forall (x,b) \in S, \quad b = 0.$$

Since $|S| < k+1$ and $\ell \geq 2$, by Lemma 19, we obtain a lower bound of $OptTE_k(c, \mathcal{M}'_n)$,

$$OptTE_k(c, \mathcal{M}'_n) = TE(c, \mathcal{M}'_n, S) \geq \frac{2^{n-\lceil \log(k+1) \rceil} + 2^{n-\ell}}{2^n}.$$

The case $k < \ell + 1$. Since $k \geq 3$, we obtain an upper bound of $OptTE_k(c, \mathcal{M}'_n)$ from Lemma 20,

$$OptTE_k(c, \mathcal{M}'_n) \leq \frac{2^{n-k+1} - 2^{n-\ell}}{2^n} < \frac{2^{n-\lceil \log(k+1) \rceil} + 2^{n-\ell}}{2^n}.$$

This is a contradiction.

The case $k = \ell + 1$. We also obtain an upper bound of $OptTE_k(c, \mathcal{M}'_n)$,

$$OptTE_k(c, \mathcal{M}'_n) \leq \frac{2^{n-k+1} - 1}{2^n} < \frac{2^{n-\lceil \log(k+1) \rceil} + 2^{n-\ell}}{2^n}.$$

This is a contradiction. ∎

Lemma 22. *For any concept $c \in \mathcal{M}'_n$ with $\ell := var(c) \geq 2$ and integer $k \in [3, TD(c, \mathcal{M}'_n) - 1]$,*

$$OptTE_k(c, \mathcal{M}'_n) = \begin{cases} \dfrac{2^{n-k+1} - 2^{n-\ell}}{2^n} & (k < \ell + 1), \\ \dfrac{2^{n-k+1} - 1}{2^n} & (k = \ell + 1). \end{cases}$$

Proof: From Lemma 21,

$$\forall S \in OptTS_k(c, \mathcal{M}'_n),\ \exists (x,b) \in S, \quad b = 1.$$

We can identify $CONS(\{(x,1)\}, \mathcal{M}'_n)$ with $\mathcal{M}_n^+$ for the same reason in Lemma 20. Therefore the formula of $OptTE_k(c, \mathcal{M}'_n)$ is obtained by the same argument in Lemma 12. ∎

The next theorem shows the optimal teaching error of the concept class $\mathcal{M}'_n$ of monomials except the empty concept.

Theorem 23. *For any integer $k \in [1, TD(\mathcal{M}'_n) - 1]$,*

$$OptTE_k(\mathcal{M}'_n) = \frac{2^{n-k+1} - 1}{2^n}.$$

Proof: When $k \geq 3$, the formula of $OptTE_k(\mathcal{M}'_n)$ is obtained by the same argument in Theorem 13.

Let us consider the case $k = 1$. Let $c_1 := *^n$. Obviously, for any $z \in \mathcal{X}$,

$$\exists c' \in CONS(\{z\}, \mathcal{M}'_n), \quad var(c') = n.$$

Thus, we have a lower bound of $OptTE_1(c_1, \mathcal{M}'_n)$,

$$OptTE_1(c, \mathcal{M}'_n) \geq d(c_1, c') = \frac{2^n - 1}{2^n}.$$

The value is the worst case error by the definition of optimal teaching errors for $\mathcal{M}'_n$. Therefore, for any $c \in \mathcal{M}'_n$,

$$OptTE_1(c, \mathcal{M}'_n) \leq OptTE_1(c_1, \mathcal{M}'_n) = \frac{2^n - 1}{2^n}.$$

We now consider the case $k = 2$. Let $c_2 = u*^{n-1}$ such that $u \in \{0,1\}$. Since we can use the same argument for $k = \ell + 1$ in Lemma 22,

$$OptTE_2(c_2, \mathcal{M}'_n) = \frac{2^{n-1} - 1}{2^n}.$$

From Lemma 20, for any concept $c \in \mathcal{M}'_n$ with $var(c) \geq 2$, we obtain an upper bound of $OptTE_2(c, \mathcal{M}'_n)$ as

$$OptTE_2(c, \mathcal{M}'_n) \leq OptTE_2(c_2, \mathcal{M}'_n) = \frac{2^{n-1} - 1}{2^n}.$$

∎

We prove the following lemmas for the proof of the optimally incremental teachability of $\mathcal{M}'_n$.

Lemma 24. *Let $c := 1^n$ and $S_1 := \{(0^n, 0)\}$.*

$$OptTS_1(c, \mathcal{M}'_n) = \{S_1\}.$$

Proof: For any concept $c' \in CONS(S_1, \mathcal{M}'_n)$,

$$\begin{aligned} s(c,c') &\geq 0, \\ a(c,c') &= 0, \\ 0 \leq w(c,c') &\leq n-1, \text{ and} \\ w(c',c) &= 0. \end{aligned}$$

Therefore,

$$TE(c, \mathcal{M}'_n, S_1) = d(c, 1*^{n-1}) = \frac{2^{n-1} - 1}{2^n}.$$

We just have to show that for any example $(x,b) \in \mathcal{X} - S_1$,

$$TE(c, \mathcal{M}'_n, \{(x,b)\}) > TE(c, \mathcal{M}'_n, S_1).$$

When $b = 0$, since $x \neq 0^n$,

$$\exists i \in [1,n], x[i] = 1.$$

For the integer i,

$$*^{i-1}0*^{n-i} \in CONS(\{(x,b)\}, \mathcal{M}'_n).$$

Therefore,

$$\begin{aligned} TE(c, \mathcal{M}'_n, \{(x,b)\}) &\geq d(c, *^{i-1}0*^{n-i}) = \frac{2^{n-1}}{2^n} \\ &> TE(c, \mathcal{M}'_n, S_1). \end{aligned}$$

When $b = 1$,

$$*^n \in CONS(\{(x,b)\}, \mathcal{M}'_n).$$

Therefore,

$$\begin{aligned} TE(c, \mathcal{M}'_n, \{(x,b)\}) &\geq d(c, *^n) = \frac{2^n - 1}{2^n} \\ &> TE(c, \mathcal{M}'_n, S_1). \end{aligned}$$

∎

Lemma 25. *Let $c := 1^n$ and $S_{opt} := \{(1^n, 1)\} \cup \{(u, 0) \mid u \in \mathcal{N}_1(1^n)\}$.*

$$MinTS(c, \mathcal{M}'_n) = \{S_{opt}\}.$$

Proof: From Theorem 15,

$$S_{opt} \in \mathit{MinTS}(c, \mathcal{M}'_n).$$

We show that there is no other minimum teaching sets of c with regard to $\mathcal{M}'_n$. Suppose that there exists $S \in \mathit{MinTS}(c, \mathcal{M}'_n)$ such that $S \neq S_{opt}$. Let $c_i := 1^{i-1} * 1^{n-i}$ such that $i \in [1, n]$. Since $S \in \mathit{MinTS}(c, \mathcal{M}'_n)$,

$$c_i \notin \mathit{CONS}(S, \mathcal{M}'_n).$$

Let $z_i := (1^{i-1}01^{n-i}, 0)$. By the same argument in Lemma 9,

$$\{z_i \mid i \in [1, n]\} = \{(u, 0) \mid u \in \mathcal{N}_1(1^n)\} \subseteq S.$$

Since $S \in \mathit{MinTS}(c, \mathcal{M}'_n)$, $|S| = n + 1$. On the other hand, $\#\{z_i \mid i \in [1, n]\} = n$. Therefore, we must specify the target concept c by using just one example. This can be achieved only by

$$\{(1^n, 1)\} \subseteq S.$$

That is

$$S = S_{opt}.$$

This is a contradiction. ■

The next theorem disproves the optimally incremental teachability of $\mathcal{M}'_n$.

Theorem 26. *The concept class $\mathcal{M}'_n$ of monomials without the empty concept is not optimally incrementally teachable.*

Proof: The concept 1^n is not optimally incrementally teachable with respect to $\mathcal{M}'_n$, since for S_1 in Lemma 24 and S_{opt} in Lemma 25,

$$S_1 \nsubseteq S_{opt}.$$

■

4.4 Monomials

In this section, we discuss the complexity of teaching $\mathcal{M}_n$ by a restricted number of examples. For any $c \in \mathcal{M}'_n$, its teaching dimension with regard to $\mathcal{M}_n$ is the same as Theorem 15, since c_e is always inconsistent with a positive example. However, it should be noticed that

$$\mathit{TD}(c_e, \mathcal{M}_n) = 2^n,$$

since we need all negative examples to teach c_e. Therefore, the teaching dimension of $\mathcal{M}_n$ is 2^n, that is

$$\mathit{TD}(\mathcal{M}_n) = 2^n.$$

From the following fact, we can apply many useful lemmas for $\mathcal{M}'_n$ to $\mathcal{M}_n$.

Fact 27. *For any set S of examples and any concept $c \in \mathcal{M}'_n$,*

$$c \in \mathit{CONS}(S, \mathcal{M}'_n) \Rightarrow c \in \mathit{CONS}(S, \mathcal{M}_n).$$

The following lemma shows the formula calculating the error between any concept in $\mathcal{M}'_n$ and the empty concept c_e.

Lemma 28. *For any concept $c \in \mathcal{M}'_n$,*

$$d(c_e, c) = \frac{|c|}{2^n} = \frac{2^{\tilde{e}}}{2^n},$$

where $\tilde{e} := \#\{\, i \mid c[i] = * \,\}$.

Proof: Since $c_e = \emptyset$, $|c_e \Delta c| = |c| = 2^{\tilde{e}}$. The formula of $d(c_e, c)$ is obtained by dividing $2^{\tilde{e}}$ by $|X_n| = 2^n$. ■

We prove the following four lemmas for the proof of Theorem 33 which shows the optimal teaching error of $\mathcal{M}_n$.

Lemma 29. *Let $S := \{(1^n, 1)\} \cup \{(u1^{n-k+1}, 0) \mid u \in \{0, 1\}^{k-1}\}$.*

$$\mathit{TE}(c_e, \mathcal{M}_n, S) = \frac{2^{n-k+1}}{2^n}.$$

Proof: We can identify $\mathit{CONS}(\{(1^n, 1)\}, \mathcal{M}_n)$ with $\mathcal{M}^+_n$. Thus, by Lemma 8,

$$\mathit{CONS}(S, \mathcal{M}_n) = \{1^{k-1}u \mid u \in \{1, *\}^{n-k+1}\}.$$

Let $\tilde{e} := \#\{\, i \mid c[i] = * \,\}$ such that $c \in \mathit{CONS}(S, \mathcal{M}_n)$. Since $\tilde{e} \leq n - k + 1$, we obtain

$$\mathit{TE}(c_e, \mathcal{M}_n, S) = \max_{\tilde{e} \leq n-k+1} \frac{2^{\tilde{e}}}{2^n} = \frac{2^{n-k+1}}{2^n}.$$

■

Lemma 30. *For any integer $k \in [4, n]$,*

$$\forall S \in \mathit{OptTS}_k(c_e, \mathcal{M}_n), \exists (x, b) \in S, \quad b = 1.$$

Proof: Suppose that

$$\exists S \in \mathit{OptTS}_k(c_e, \mathcal{M}_n), \forall (x, b) \in S, \quad b = 0.$$

Since $|S| < k + 1$, from Lemma 17,

$$\exists c \in \mathit{CONS}(S, \mathcal{M}_n), \quad var(c) = \lceil \log(k + 1) \rceil.$$

For the concept c, we have a lower bound of $\mathit{OptTS}_k(c_e, \mathcal{M}_n)$ as

$$\mathit{OptTS}_k(c_e, \mathcal{M}_n) \geq d(c_e, c) = \frac{2^{n - \lceil \log(k+1) \rceil}}{2^n}.$$

However, we have an upper bound of $\mathit{OptTS}_k(c_e, \mathcal{M}_n)$ in Lemma 29 as

$$\mathit{OptTS}_k(c_e, \mathcal{M}_n) \leq \frac{2^{n-k+1}}{2^n} < \frac{2^{n - \lceil \log(k+1) \rceil}}{2^n},$$

since $k \geq 4$. This is a contradiction. ■

Lemma 31. *For any set S of examples with $\mathit{CONS}(S, \mathcal{M}_n) \neq \emptyset$,*

$$\exists (x, 1) \in S \quad \Rightarrow \quad \mathit{TE}(c_e, \mathcal{M}_n, S) \geq \frac{2^{n-k+1}}{2^n},$$

where $k := |S|$.

Proof: We can identify $\mathit{CONS}(\{(x, 1)\}, \mathcal{M}_n)$ with $\mathcal{M}^+_n$. Let $S^- := \{(x', b) \in S \mid b = 0\}$. Obviously $|S^-| < k$. From Lemma 9,

$$\exists c \in \mathit{CONS}(S, \mathcal{M}_n), \quad var(c) < k,$$

Let $\ell := var(c)$. Since $\ell \leq k - 1 < k$,

$$\mathit{TE}(c_e, \mathcal{M}_n, S) \geq d(c_e, c) \geq \frac{2^{n-\ell}}{2^n} \geq \frac{2^{n-k+1}}{2^n}.$$

■

Lemma 32. *For any* $k \in [1, TD(\mathcal{M}_n) - 1]$,

$$OptTE_k(c_e, \mathcal{M}_n) = \begin{cases} \frac{2^{n-k}}{2^n} & (k \le 2), \\ \frac{2^{n-k+1}}{2^n} & (2 < k \le n), \\ \frac{1}{2^n} & (n < k). \end{cases}$$

Proof: We first consider the case $n < k$. Since we can specify a concept 1^n with only $n + 1$ examples by using S_ℓ ($\ell = n$) in Theorem 15, we obtain an upper bound of $OptTE_k(c_e, \mathcal{M}_n)$ as

$$OptTE_k(c_e, \mathcal{M}_n) \le TE(c_e, \mathcal{M}_n, S_\ell) = d(c_e, 1^n) = \frac{1}{2^n}.$$

Since we cannot specify c_e with less than 2^n examples, we obtain an lower bound of $OptTE_k(c_e, \mathcal{M}_n)$ as

$$OptTE_k(c_e, \mathcal{M}_n) \ge \frac{1}{2^n} > \frac{0}{2^n}.$$

Therefore, we obtain the formula for $n < k$.

Next we consider the case $4 \le k \le n$. For any $S \in OptTS_k(c_e, \mathcal{M}_n)$, S has a positive example by using Lemma 30. Thus, we obtain a lower bound of $OptTE_k(c_e, \mathcal{M}_n)$ from Lemma 31,

$$OptTE_k(c_e, \mathcal{M}_n) \ge \frac{2^{n-k+1}}{2^n}.$$

By using the set S of examples in Lemma 29, we obtain an upper bound of $OptTE_k(c_e, \mathcal{M}_n)$,

$$OptTE_k(c_e, \mathcal{M}_n) \le TE(c_e, \mathcal{M}_n, S) = \frac{2^{n-k+1}}{2^n}.$$

Therefore, we obtain the formula for $4 \le k \le n$.

We now consider the case $k = 3$. For any set $S \in OptTS_3(c_e, \mathcal{M}_n)$, if S does not contain a positive example, then we obtain a lower bound of $OptTE_3(c_e, \mathcal{M}_n)$ by using Lemma 17,

$$OptTE_k(c_e, \mathcal{M}_n) = \frac{2^{n-\lceil \log 4 \rceil}}{2^n} = \frac{2^{n-2}}{2^n}.$$

We obtain the formula for $k = 3$ by using the same argument for $4 \le k \le n$.

Moreover, we consider the case $k = 2$. Let us consider the set $S_2 := \{(1^n, 0), (0^n, 0)\}$ of examples. The set S_2 gives an upper bound of $OptTE_2(c_e, \mathcal{M}_n)$ as

$$OptTE_2(c_e, \mathcal{M}_n) \le TE(c_e, \mathcal{M}_n, S_2) = d(c_e, 10{*}^{n-2}) = \frac{2^{n-2}}{2^n}.$$

For any $S \in OptTS_2(c_e, \mathcal{M}_n)$, S does not contain a positive example, because if S contains a positive example then a lower bound of $TE(c_e, \mathcal{M}_n, S)$ calculated by Lemma 31 is greater than the upper bound. By using Lemma 17, we obtain a lower bound of $OptTE_2(c_e, \mathcal{M}_n)$ as

$$OptTE_2(c_e, \mathcal{M}_n) \ge \frac{2^{n-\lceil \log 3 \rceil}}{2^n} \ge \frac{2^{n-2}}{2^n}.$$

Therefore, we obtain the formula for $k = 2$.

Finally we consider the case $k = 1$. Let $S_1 := \{(1^n, 0)\}$. We obtain the formula for $k = 1$ by using the same argument for $k = 2$ with regard to S_1. ■

The next theorem shows the optimal teaching error of the concept class $\mathcal{M}_n$ of monomials.

Theorem 33. *For any* $k \in [1, TD(\mathcal{M}_n) - 1]$,

$$OptTE_k(\mathcal{M}_n) = \begin{cases} \frac{2^{n-k+1} - 1}{2^n} & (k \le 2), \\ \frac{2^{n-k+1}}{2^n} & (2 < k \le n), \\ \frac{1}{2^n} & (n < k). \end{cases}$$

Proof: The formula is obtained by taking the worst case in Lemma 23 and Lemma 32. ■

Theorem 34. *The concept class $\mathcal{M}_n$ of monomials is not optimally incrementally teachable.*

Proof: The concept 1^n is also not optimally incrementally teachable with respect to $\mathcal{M}_n$. ■

Finally, we consider some properties about each concept in $\mathcal{M}_n$.

The next theorem shows that the concept $*^n$, which is the most easy to teach in the classical model, can be the most difficult to teach in our model. At the same time, the theorem shows that the empty concept c_e, which is the most difficult to teach in the classical model, can be relatively easy to teach in our model.

Theorem 35. *For any concept* $c \in \mathcal{M}_n - \{*^n\}$,

$$OptTE_1(c, \mathcal{M}_n) < OptTE_1(*^n, \mathcal{M}_n).$$

Proof: $OptTE_1(*^n, \mathcal{M}_n)$ is obtained in the case $k = 1$ in Lemma 23. For any concept $c \in \mathcal{M}_n - \{*^n\}$, an upper bounds of $OptTE_1(c, \mathcal{M}_n)$ is calculated by Lemma 20 and Lemma 32. ■

The next theorem shows that when the number k of examples is restricted to $4 \le k < 2^{n-1}$, the teacher must use inconsistent examples for *optimally incrementally* teaching $\mathcal{M}_n$. It should be noted that the theorem derives Theorem 3 and Theorem 4. The natural concept class $\mathcal{M}_n$ also has the same properties as the simple self-made concept classes used in the proofs of those theorems.

Theorem 36. *For any integer* $k \in [4, 2^{n-1} - 1]$,

$$\forall S \in OptTS_k(c_e, \mathcal{M}_n), \quad c_e \notin CONS(S, \mathcal{M}_n).$$

Proof: When $4 \le k \le n$, it is clear from Lemma 30.

For the case $n + 1 \le k \le 2^{n-1} - 1$, we suppose that

$$\exists S \in OptTS_k(c_e, \mathcal{M}_n), \quad c_e \in CONS(S, \mathcal{M}_n).$$

Since $c_e \in CONS(S, \mathcal{M}_n)$, S contains negative examples only. Obviously $|S| < k + 1$. From Lemma 17,

$$\exists c_k \in CONS(S, \mathcal{M}_n), \quad var(c_k) = \lceil \log(k + 1) \rceil.$$

Therefore, we obtain a lower bound of $OptTE_k(c_e, \mathcal{M}_n)$ as

$$OptTE_k(c_e, \mathcal{M}_n) \ge d(c_e, c_k) \ge \frac{2^{n-\lceil \log(k+1) \rceil}}{2^n} \ge \frac{2}{2^n}.$$

However, the set S_ℓ ($\ell = n$) in Theorem 15 can specify a concept c' with $var(c') = n$, that is, we obtain an upper bound of $OptTE_k(c_e, \mathcal{M}_n)$ as

$$OptTE_k(c_e, \mathcal{M}_n) = TE(c_e, \mathcal{M}_n, S_\ell) = d(c_e, c') = \frac{1}{2^n}.$$

This is a contradiction. ■

Table 3: Summary. The last two rows (†) are newly proved in this paper.

$\mathcal{C}$	$\mathcal{M}_n^+$	$\mathcal{M}'_n$	$\mathcal{M}_n$
$TD(\mathcal{C})$	n	$n+1$	2^n
Teachability	True	True	False
$OptTE_k(\mathcal{C})^\dagger$	$\frac{2^{n-k}-1}{2^n}$	$\frac{2^{n-k+1}-1}{2^n}$	$\begin{cases} \frac{2^{n-k+1}-1}{2^n} & (k \leq 2) \\ \frac{2^{n-k+1}}{2^n} & (2 < k \leq n) \\ \frac{1}{2^n} & (n < k) \end{cases}$
Optimally Incremental Teachability†	True	False	False

5 Conclusions

In this paper, we studied the complexity of teaching in the situation that the number of examples is restricted. We formulated a new model of teaching. In our model, we measure the complexity of a concept class by its teaching error to a target concept in the worst case. We say that a concept class is optimally incrementally teachable if the teacher can optimally teach it to the learner whenever teaching is terminated.

Table 3 summarizes the new results together with the previously known results in terms of the three concept classes of monotone monomials $\mathcal{M}_n^+$, monomials without the empty concept $\mathcal{M}'_n$, and monomials $\mathcal{M}_n$. $TD(\mathcal{C})$ means the teaching dimension of a concept class $\mathcal{C}$, and $OptTE_k(\mathcal{C})$ means the optimal teaching error of a concept class $\mathcal{C}$ by at most k examples in our model. The table indicates that $OptTE_k(\mathcal{C})$ becomes greater as $TD(\mathcal{C})$ becomes greater, regardless of a restriction k. However, the boundary of optimally incremental teachability is different from that of polynomial teachability in the classical model. It should be noted that the boundary lies on natural concept classes that are well known and studied in computational learning theory.

Focusing on each concept in $\mathcal{M}_n$, we showed that some concepts are easy to teach in our new model although they are difficult to teach in the classical model, and vice versa. We also showed that the teacher must *tell lies* (i.e. use inconsistent examples) to teach the empty concept in $\mathcal{M}_n$ in order to avoid a serious misunderstanding in short words. This result corresponds to the heuristics that when humans teach something within a time limit, they should over-simplify it by ignoring a few exceptions.

Finally, we suggest the following open problems. Are there two natural concept classes $\mathcal{C}_1$ and $\mathcal{C}_2$ such that for some integer k, $OptTE_k(\mathcal{C}_1) < OptTE_k(\mathcal{C}_2)$ despite $TD(\mathcal{C}_1) > TD(\mathcal{C}_2)$? What if the model is extended to infinite instance spaces and allows all other distributions? What if we consider the situation that the number of examples is restricted in PAC model? How does the model relate to the other models?

References

[Ang88] Dana Angluin. Queries and Concept Learning. *Machine Learning*, 2(4):319–342, 1988.

[Bal08] Frank J. Balbach. Measuring teachability using variants of the teaching dimension. *Theoretical Computer Science*, 397(1–3):94–113, 2008.

[BZ06] Frank J. Balbach and Thomas Zeugmann. Teaching Randomized learners. In *Proceedings of the 19th Annual Conference on Learning Theory (COLT 2006)*, volume 4005 of *LNCS*, pages 229–243. Springer-Verlag, 2006.

[GK95] Sally A. Goldman and Michael J. Kearns. On the Complexity of Teaching. *Journal of Computer and System Sciences*, 50:303–314, 1995.

[Gol67] E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[Han07] Steve Hanneke. Teaching Dimension and the Complexity of Active Learning. In *Proceedings of the 20th Annual Conference on Learning Theory (COLT 2007)*, volume 4539 of *LNCS*, pages 66–81. Springer-Verlag, 2007.

[Heg95] Tibor Hegedűs. Generalized Teaching Dimensions and the Query Complexity of Learning. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory (COLT 1995)*, pages 108–117. ACM Press, 1995.

[JT92] Jeffrey Jackson and Andrew Tomkins. A Computational Model of Teaching. In *Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory (COLT 1992)*, pages 319–326. ACM Press, 1992.

[LSW07] Homin K. Lee, Rocco A. Servedio, and Andrew Wan. DNF are teachable in the average case. *Machine Learning*, 69(2–3):79–96, 2007.

[Mat97] H. David Mathias. A Model of Interactive Teaching. *Journal of Computer and System Sciences*, 54(3):487–501, 1997.

[SM91] Ayumi Shinohara and Satoru Miyano. Teachability in Computational Learning. *New Generation Computing*, 8(4):337–347, 1991.

[Val84] Leslie G. Valiant. A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[ZLHZ08] Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Teaching Dimensions Based on Cooperative Learning. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, pages 135–146. Omnipress, 2008.