How Many Queries are Needed to Learn One Bit of Information?

Hans-Ulrich Simon
Fakultät für Informatik
Lehrstuhl Mathematik und Informatik
Ruhr-Universitaet Bochum
D-44780 Bochum
simon@lmi.ruhr-uni-bochum.de

December 18, 2002

Abstract

In this paper we study the question how many queries are needed to "halve a given version space". In other words: how many queries are needed to extract from the learning environment the one bit of information that rules out fifty percent of the concepts which are still candidates for the unknown target concept. We relate this problem to the classical exact learning problem. For instance, we show that lower bounds on the number of queries needed to halve a version space also apply to randomized learners (whereas the classical adversary arguments do not readily apply). Furthermore, we introduce two new combinatorial parameters, the halving dimension and the strong halving dimension, which determine the halving complexity (modulo a small constant factor) for two popular models of query learning: learning by a minimum adequate teacher (equivalence queries combined with membership queries) and learning by counterexamples (equivalence queries alone). These parameters are finally used to characterize the additional power provided by membership queries (compared to the power of equivalence queries alone). All investigations are purely information-theoretic and ignore computational issues.

1 Introduction

The exact learning model was introduced by Angluin in [1]. In this model, a learner A tries to identify an unknown target concept $C_* \in \mathcal{C}$ (of the form $C_* : X \rightarrow \{0,1\}$ for a finite set X) by means of queries, where $\mathcal{C}$ denotes the class of target concepts (known to the learner). A aims at identifying C_* with as few queries as possible. A is exposed to a worstcase analysis: we can think of a malicious adversary ADV who picks the target concept C_* which is hardest-to-learn for A and selects its answers to queries such as to hide the identity of C_* as long as possible. The progress made by A during learning can be measured in terms of the current "version space" $\mathcal{V}$. $\mathcal{V}$ contains all concepts from $\mathcal{C}$ that are still conceivable (given the answers received by A so far). Initially, $\mathcal{V} = \mathcal{C}$. During learning $\mathcal{V}$ shrinks until it finally coincides with $\{C_*\}$. From this perspective, A aims at shrinking $\mathcal{V}$ as fast as pssible, whereas the adversary tries to slow down this process. Note, however, that the answers of ADV must be "honest", which means that C_* (the target concept selected by ADV in the beginning) must always belong to $\mathcal{V}$. The initial commitment to C_* is binding in this sense.

It is well-known in the learning community that, as far as deterministic learners are concerned, the initial commitment to C_* is not essential. Lower bounds on the number of queries needed to identify a target concept from $\mathcal{C}$ are typically shown by an adversary argument, where there is no initial commitment to a target concept. The answers of ADV are considered as honest as long as the current version space $\mathcal{V}$ is not empty. The learner must proceed until $\mathcal{V}$ is a singleton set. At first glance, the setting without an initial commitment to a target concept seems to give more power to the adversary. However, if we assume that A is deterministic, both settings require the same number of queries: since A is deterministic, one can "predict" which concept C_* will form the final (singleton) version space. Now, it does not hurt the adversary, to commit itself to C_* as the target concept in the beginning.

In order to distinguish the two settings, we speak of a "learning game" when there is an initial commitment and of a "contraction game" otherwise. As for deterministic learners, there is not much of a difference between these games. However, as we will see in the course of this paper, the two settings *are* essentially different when we consider randomized learners. Thus, typical and widely used adversary arguments cannot be readily applied to determine (or lower-bounding) randomized learning complexity.[1]

The main issue in this paper is the number of queries needed to halve (as opposed to contract) an initial version space $\mathcal{V}_0$. There are several reasons for this kind of research:

- Contraction of the version space by iterated halving is considered as very efficient. Iterated halving is an important building block of some well known strategies such as the "Majority Vote Strategy". The binary search paradigm is based on halving. Halving may therefore be considered as an interesting problem in its own right.

- Halving the version space yields exactly one bit of information. In this sense, we explore the hardness to extract one bit of information from the learning environment. This sounds like an elementary and natural problem.

[1] To the best of our knowledge, almost all papers devoted to exact learning assume deterministic learning algorithms. A notable exception is the paper [4] of Maass that demonstrates the significance of supporting examples in the on-line learning model, when the learner is randomized and the learning environment is oblivious.

- Although the contraction game does not readily lead to lower bounds on randomized learning complexity, the halving game fits precisely this purpose.
- We can characterize the halving complexity for two popular query types (equivalence and membership queries) by tight combinatorial bounds (leaving almost no gap).[2] These bounds can be used to characterize the additional power provided by membership queries (compared to the power of equivalence queries alone).

The paper is organized as follows. In Section 2, we present the basic definitions and notations. In particular, we introduce the learning game, the contraction game and the halving game. In Section 3, we investigate the relation between halving and learning complexity (including randomized learning complexity). In Section 4, we present combinatorial (lower and upper) bounds on the halving complexity. In Section 5, these bounds are used to characterize the additional power provided by membership queries (compared to the power of equivalence queries alone).

2 Basic Definitions and Notations

Let X be a finite set and $\mathcal{C}, \mathcal{H}$ be two families of functions from X to $\{0,1\}$. In the sequel, we refer to X as the *instance space*, to $\mathcal{C}$ as the *concept class*, and to $\mathcal{H}$ as the *hypothesis class*. It is assumed that $\mathcal{C} \subseteq \mathcal{H}$. A labeled instance $(x,b) \in X \times \{0,1\}$ is called a *sample-point*. A *sample* is a collection of sample-points. For convenience, we represent each sample S as a partially defined binary function over domain X. More formally, S is of the form $S : X \rightarrow \{0,1,?\}$, where $S(x) =?$ indicates that S is undefined on instance x. The set

$$\mathrm{supp}(S) := \{x \in X : S(x) \neq ?\} \tag{1}$$

is called the *support* of S. Note that a concept or hypothesis can be viewed as a sample with full support. The *size* of S is the number of instances in its support. S' is called *subsample* of S, denoted as $S' \sqsubseteq S$, if $\mathrm{supp}(S') \subseteq \mathrm{supp}(S)$ and $S'(x) = S(x)$ for each instance $x \in \mathrm{supp}(S')$. We say that sample S and concept C are *consistent* if $S \sqsubseteq C$. We say that S *has a consistent explanation in* $\mathcal{C}$ if there exists a concept $C \in \mathcal{C}$ such that S and C are consistent. The terminology for hypotheses is analogous.

In the exact learning model, a *learner (learning algorithm)* A has to identify an unknown *target concept* $C_* \in \mathcal{C}$ by means of queries that must be honestly answered by an oracle. Among the most popular query types are the following ones:

[2] The derivation of these bounds is based on ideas and results from [3, 2], where bounds on the number of queries needed for learning (i.e, for contracting the version space) are presented. These bounds, however, leave a gap. It seems that the most accurate combinatorial bounds are found on the level of the halving task.

Equivalence Queries A selects a hypothesis H from its hypothesis class $\mathcal{H}$. If $H = C_*$, the oracle answers "YES" (signifying that A suceeded to identify the target concept). Otherwise, the oracle returns a counterexample, i.e., an $x \in X$ such that $H(x) \neq C_*(x)$.

Membership Queries A selects an $x \in X$ and receives label $C_*(x)$ from the oracle.

In the sequel, $\mathcal{Q}$ denotes the allowed types of queries. Given a collection of queries and answers, the current *version space* $\mathcal{V}$ is the set of concepts from $\mathcal{C}$ that are still conceivable. $\mathcal{V}[Q,R]$ denotes the version space resulting from the current version space $\mathcal{V}$ when A issues query Q and receives answer R.

Example 2.1 *Let $\mathcal{V}$ be the current version space. If A issues a membership query with instance x and receives the binary label b, then the subsequent version space is given by*

$$\mathcal{V}[x,b] := \{C \in \mathcal{V} : C(x) = b\}. \tag{2}$$

Similarly, if A issues an equivalence query with hypothesis H and receives the counterexample x, then the subsequent version space is given by

$$\mathcal{V}[H,x] := \{C \in \mathcal{V} : C(x) \neq H(x)\}. \tag{3}$$

Clearly, answer "YES" to an equivalence query immediately leads to the final version space $\{C_*\}$.

Since the exact learning model exposes the learner to a worstcase analysis, it will be technically convenient to cast exact learning as a game between the learner A and a malicious adversary ADV. A aims at contracting the version space to a singleton (containing the target concept) as fast as possible. ADV tries to slow down the learning process as much as possible. Details follow.

The Learning Game $\mathcal{C}, \mathcal{H}$ and $\mathcal{Q}$ are fixed and known to both players. The game proceeds as follows. In a first move (invisible to A), ADV picks the target concept C_* from C. Afterwards, both players proceed in rounds. In each round, first player A makes its move by selecting a query from $\mathcal{Q}$. Then, ADV makes its move by selecting an honest answer. The answer is considered as honest if the target concept belongs to the resulting version space. The game terminates when the current version space coincides with $\{C_*\}$.

Deterministic Learning Complexity Given two deterministic players A and ADV, we define $\mathrm{DLC}^{\mathcal{Q}}_{A,ADV}(\mathcal{C},\mathcal{H})$ as the number of rounds in the learning game until termination. The *deterministic learning complexity* associated with $\mathcal{C}, \mathcal{H}, \mathcal{Q}$ is then given by

$$\mathrm{DLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) := \min_{A} \max_{ADV} \mathrm{DLC}^{\mathcal{Q}}_{A,ADV}(\mathcal{C},\mathcal{H}) \ , \tag{4}$$

where minimization and maximization range over all deterministic playing strategies.

In the classical model of exact learning, it is usually assumed that the learner A is *deterministic*. Thus, our notion of $\mathrm{DLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$ coincides with the classical definition of learning complexity.

Randomized Learning Complexity Given two randomized players A and ADV[3], we define $\mathrm{RLC}^{\mathcal{Q}}_{A,ADV}(\mathcal{C},\mathcal{H})$ as the expected number of rounds in the learning game until termination. The *randomized learning complexity* associated with $\mathcal{C},\mathcal{H},\mathcal{Q}$ is then given by

$$\mathrm{RLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) := \min_{A} \max_{ADV} \mathrm{RLC}^{\mathcal{Q}}_{A,ADV}(\mathcal{C},\mathcal{H}) \ , \qquad (5)$$

where minimization and maximization range over all randomized playing strategies. The definition does however not change when maximization ranges only over all deterministic strategies, because among the strongest adversaries of A there is always a deterministic one (which can be shown by a standard argument).

Lower bounds on the deterministic learning complexity are typically shown by adversary arguments, where the adversary does *not* make an initial commitment to a target concept. We associate the following game with this kind of interaction between the learner and the adversary.

The Contraction Game $\mathcal{C},\mathcal{H}$ and $\mathcal{Q}$ are fixed and known to both players. The game proceeds in rounds. In each round, first player A makes its move by selecting a query from $\mathcal{Q}$. Then, ADV makes its move by selecting an honest answer. Here, the answer is considered as honest if the resulting version space is not empty. The game terminates when the current version space consists of a single concept.

The following lemmas (given without proof) recall some well-known facts (in a slightly more general setting).

Lemma 2.2 *As for the contraction game, there exist two deterministic optimal players A_* and ADV_*, i.e., the following holds:*

1. *Let A be any (possibly randomized) contraction algorithm. Then, ADV_* forces A to make at least as many moves as A_*.*
2. *Let ADV be any (possibly randomized) oracle. Then A_* needs no more moves against ADV than against ADV_*.*

Lemma 2.2 implies that A_* is the best contraction algorithm among all (possibly randomized) algorithms. The lemma implies also that ADV_* is the strongest adversary of A_*.

[3]Here, ADV knows the program of A, but A determines its next move by means of secret random bits. Thus, ADV_A knows the probability distribution of the future moves, but cannot exactly predict them. This corresponds to what is called "weakly oblivious environment" in [4].

Lemma 2.3 *$DLC^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$ coincides with the number of rounds in the contraction game between A_* and ADV_*.*

Lemma 2.3 allows to analyze the deterministic learning complexity by means of the conceptually simpler contraction game. The following example however shows that Lemma 2.3 cannot be generalized from deterministic to randomized learners.

Example 2.4 *Consider the model of learning by means of equivalence queries. Let the concept and hypothesis class coincide with the set of all functions from X to $\{0,1\}$, where $X = \{1,\ldots,d\}$. Clearly, the contraction task forces each deterministic (or randomized) algorithm to issue d equivalence queries because each (non-redundant) query halves the version space. As for deterministic learners, the same remark is valid for the learning game. However the following randomized learner needs only $d/2$ queries on the average:*
Pick a first hypothesis $H_0 : X \to \{0,1\}$ uniformly at random. Given that the current hypothesis is H and that counterexample $x \in X$ is received, let the next hypothesis H' coincide with H on $X \setminus \{x\}$ and set $H'(x) := 1 - H(x)$.
The number of rounds in a learning game with target concept C_* equals the number of instances on which C_* and H_0 disagree. This is $d/2$ on the average.

Example 2.4 shows that the contraction game does not lead to lower bounds on the randomized learning complexity (at least not in the simple way as it does for deterministic learners). The following game will better fit this purpose.

The Halving Game The halving game is defined like the contraction game except that it may start with an arbitrary initial version space $\mathcal{V}_0 \subseteq \mathcal{C}$ (known to both players), and it terminates as soon as the current version space $\mathcal{V}$ contains at most half of the concepts of $\mathcal{V}_0$. Player A is called *halving algorithm.*

Halving Complexity Like in the contraction game, there exist two optimal deterministic players: A_* (representing the optimal halving algorithm) and ADV_* (wich is also the strongest adversary for A_*). (Compare with Lemma 2.2.) Let $\mathrm{HC}^{\mathcal{Q}}(\mathcal{V}_0,\mathcal{H})$ be defined as the number of rounds in the halving game between A_* and ADV_*. This parameter has the disadvantage of being not monotonous: a subset of $\mathcal{V}_0$ might be harder to halve than $\mathcal{V}_0$ itself. In order to force monotonicity, we define the *halving complexity* associated with $\mathcal{C},\mathcal{H},\mathcal{Q}$ as follows:

$$\mathrm{HC}_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) = \max\{\mathrm{HC}^{\mathcal{Q}}(\mathcal{V},\mathcal{H}) : \mathcal{V} \subseteq \mathcal{C}\} \tag{6}$$

The relation between halving and learning is the central issue of the next section.

3 Halving and Learning Complexity

Theorem 3.1 *The following chain of inequalities is valid:*

$$\frac{1}{2} \cdot HC_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) \quad < \quad RLC^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) \tag{7}$$

$$
\begin{aligned}
&\leq \max\{HC_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H}), RLC^{\mathcal{Q}}(\mathcal{C},\mathcal{H})\} && (8)\\
&\leq DLC^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) && (9)\\
&\leq \lfloor \log|\mathcal{C}| \rfloor \cdot HC_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H}). && (10)
\end{aligned}
$$

Proof

Proof of Inequality (7) Let $\mathcal{V}_* \subseteq \mathcal{C}$ be the version space such that

$$q_* := \mathrm{HC}_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) = \mathrm{HC}^{\mathcal{Q}}(\mathcal{V}_*,\mathcal{H}).$$

Let ADV_* be the optimal deterministic adversary for the problem of halving $\mathcal{V}_*$. Thus, ADV_* forces each halving algorithm to issue at least q_* queries. We have to show that $\mathrm{RLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) > q_*/2$.
The quantity $\mathrm{RLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$ is defined as the expected number of rounds in the learning game between the best randomized learner A and its strongest (deterministic) adversary ADV_A. In order to show that $\mathrm{RLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) > q_*/2$, we will modify the learning game and we will use a randomized (and possibly suboptimal) adversary ADV. The modifications of the learning game will be in favour of the learner. Therefore, it suffices to show that the expected number of rounds in the modified game between the best randomized player A of the modified game and ADV exceeds $q_*/2$. Note that, since ADV is randomized, the expectation must be taken with respect to the random decisions of A *and* ADV.
The modifications of the learning game will make it more similar to the halving game. They are as follows:

1. The target concept C_* must be chosen from $\mathcal{V}_*$. Both players know that $\mathcal{V}_*$ is the initial version space.

2. The game stops when the adversary gives up (early stopping condition) or when the current version space contains at most half of the concepts of $\mathcal{V}_*$ (regular stopping condition).

Let A be the strongest randomized player of the modified game. Its adversary ADV operates as follows:

First move ADV picks a target concept $C_* \in \mathcal{V}_*$ uniformly at random.

Subsequent moves ADV maintains a simulation of ADV_*. However, if an answer of ADV_* excludes the target concept C_* from the resulting version space, then ADV gives up (early stopping condition).

What happens if we substitute ADV_* for ADV in the modified game? First, A has no way to distinguish between a game played against ADV_* and a game played against ADV that reached the regular stopping condition. We may actually view ADV_* as an unfair version of ADV that does not give up when the early stopping condition is reached. Second, the modified game between A and ADV_* is nothing but the halving game. Thus, it will always take at least q_*

rounds.
The proof of inequality (7) can now be completed as follows. Let $Z_* \geq q_*$ be the random variable that counts the number of rounds in the modified game between A and ADV_* (where ADV_* is viewed as the unfair version of ADV). Let $Z \leq Z_*$ be the random variable that counts the number of rounds in the learning game between A and ADV, or equivalently, the number of rounds in the modified game between A and ADV_* until one of the stopping conditions (including the early stopping condition) is reached. Note that $Z < Z_*$ if and only if the regular stopping condition is not reached. We claim that the probability of this to happen is less than 1/2. From the claim, the lower bound is obtained as follows:

$$E[Z] \geq E[Z|Z = Z_*]\,\Pr[Z = Z_*] = E[Z_*|Z = Z_*]\,\Pr[Z = Z_*] > \frac{q_*}{2} \ .$$

We still have to show that the probability of reaching the regular stopping condition exceeds 1/2. Let q be the number of rounds in the game between A and ADV_*. It follows that $\mathcal{V}_{q-1}$ (the version space before the final round) still contains more than half of the concepts in $\mathcal{V}_*$. Thus, with a probability exceeding 1/2, the target concept C_* was taken from $\mathcal{V}_{q-1}$.[4] Thus, with probability exceeding 1/2, the regular stopping condidion is reached.

Proof of Inequality (9) Deterministic query learning is correctly modelled by the contraction game. Since the contraction of an initial version space $\mathcal{V}_*$ cannot be easier than halving $\mathcal{V}_*$, we get $\mathrm{DLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) \geq \mathrm{HC}_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$. Clearly, $\mathrm{DLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H}) \geq \mathrm{RLC}^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$. Thus, Inequality (9) is valid.

Proof of Inequality (10) Contraction of the initial version space $\mathcal{C}$ is obtained after $\lfloor \log|\mathcal{C}| \rfloor$ halvings. A learning algorithm A may therefore proceed in $\lfloor \log|\mathcal{C}| \rfloor$ phases and apply, in each phase with current version space $\mathcal{V}$, the optimal deterministic halving strategy w.r.t. initial version space $\mathcal{V}$. In each phase, at most $\mathrm{HC}_*^{\mathcal{Q}}(\mathcal{C},\mathcal{H})$ queries are issued.
Since Inequality (8) is trivial, we are done. •

4 Bounds on the Halving Complexity

For $\mathcal{Q} = \{\mathrm{EQ},\mathrm{MQ}\}$ or $\mathcal{Q} = \{\mathrm{EQ}\}$, the halving complexity can be characterized by combinatorial parameters that we call "halving dimension" and "strong halving dimension".[5] These parameters are closely related to the "consistency dimension" $\mathrm{cdim}(\mathcal{C},\mathcal{H})$ and "strong consistency dimension" $\mathrm{scdim}(\mathcal{C},\mathcal{H})$ that

[4] Here, we use the fact that the random decisions of A and ADV_* and the random choice of C_* in the first move of ADV are statistically independent.
[5] Note that the case $\mathcal{Q} = \{\mathrm{EQ},\mathrm{MQ}\}$ and $\mathcal{H} = \emptyset$ covers also the case $\mathcal{Q} = \{\mathrm{MQ}\}$.

were used in [2] for bounding the deterministic learning complexity from below and above as follows[6]:

$$\begin{array}{rcccl} \mathrm{cdim}(\mathcal{C},\mathcal{H}) & \le & \mathrm{DLC}^{EQ,MQ}(\mathcal{C},\mathcal{H}) & \le & 1+\lceil \log|\mathcal{C}|\cdot \mathrm{cdim}(\mathcal{C},\mathcal{H})\rceil \\ \mathrm{scdim}(\mathcal{C},\mathcal{H}) & \le & \mathrm{DLC}^{EQ}(\mathcal{C},\mathcal{H}) & \le & 1+\lceil \ln|\mathcal{C}|\cdot \mathrm{scdim}(\mathcal{C},\mathcal{H})\rceil \end{array}$$

These bounds leave a gap of size $\theta(\log|\mathcal{C}|)$. The bounds that we will derive in the course of this section are (almost) tight.

Definition 4.1 *1. The parameter $Hdim(\mathcal{V},\mathcal{H})$ denotes the smallest number d with the following property. If a function $S: X \to \{0,1\}$ (sample with full support) does not belong to $\mathcal{H}$, then there exists a a subsample $S' \sqsubseteq S$ of size at most d such that the fraction of concepts in $\mathcal{V}$ that are consistent with S' is at most $1/2$. The* halving dimension *associated with $\mathcal{C}$ and $\mathcal{H}$ is given by*

$$Hdim_*(\mathcal{C},\mathcal{H}) := \max\{Hdim(\mathcal{V},\mathcal{H}) : \mathcal{V} \subseteq \mathcal{C}\}. \tag{11}$$

2. The parameter $SHdim(\mathcal{V},\mathcal{H})$ denotes the smallest number d with the following property. If a sample $S: X \to \{0,1,?\}$ has no consistent explanation in $\mathcal{H}$, then there exists a a subsample $S' \sqsubseteq S$ of size at most d such that the fraction of concepts in $\mathcal{V}$ that are consistent with S' is at most $1/2$. The strong halving dimension *associated with $\mathcal{C},\mathcal{H},\mathcal{Q}$ is given by*

$$SHdim_*(\mathcal{C},\mathcal{H}) := \max\{SHdim(\mathcal{V},\mathcal{H}) : \mathcal{V} \subseteq \mathcal{C}\}. \tag{12}$$

Note that both definitions are almost identical, except for the subtle fact that the first definition ranges over samples with full support, whereas the second definition ranges over all samples. The next theorems show that the halving dimension characterizes the halving complexity when $\mathcal{Q} = \{\mathrm{EQ},\mathrm{MQ}\}$, and the strong halving dimension characterizes the halving complexity when $\mathcal{Q} = \{\mathrm{EQ}\}$.

Theorem 4.2 *$HC^{EQ,MQ}(\mathcal{V},\mathcal{H}) = Hdim(\mathcal{V},\mathcal{H})$.*

Proof For sake of simplicity, set $q := \mathrm{HC}^{EQ,MQ}(\mathcal{V},\mathcal{H})$ and $d := \mathrm{Hdim}(\mathcal{V},\mathcal{H})$.

First, we show that $q \ge d$. The minimality of d implies that there exists a function $C: X \to \{0,1\} \notin \mathcal{H}$ such that each subsample $S \sqsubseteq C$ of size at most $d-1$ is consistent with more than half of the concepts in $\mathcal{V}$. Thus, any halving algorithm issuing equivalence queries with hypotheses from $\mathcal{H}$ fails to be consistent with C and may obtain counterexamples taken from C. If, in addition, each membership query is answered in accordance with C, then the sample points, returned after up to $d-1$ queries, form a subsample $S \sqsubseteq C$ of size at most $d-1$. Hence, at least one additional query is needed for halving the initial version space $\mathcal{V}$.

Second, we show that $q \le d$. Let $M: X \to \{0,1\}$ be a function that goes with the majority of the concepts in $\mathcal{V}$ on each instance $x \in X$ (breaking

[6] As for the formal definition of $\mathrm{cdim}(\mathcal{C},\mathcal{H})$ and $\mathrm{scdim}(\mathcal{C},\mathcal{H})$, see [2].

ties arbitrarily). If $M \in \mathcal{H}$, then $q = 1$ because the equivalence query with hypothesis M will halve the version space $\mathcal{V}$. Clearly, $d \geq 1$. We may therefore assume wlog. that $M \notin \mathcal{H}$. It follows that there exists a subsample $S \sqsubseteq M$ of size at most d that is inconsistent to at least half of the concepts in $\mathcal{V}$. The crucial observation is that $q \leq |S| \leq d$ since $\mathcal{V}$ can be halved by issuing the $|S|$ membership queries for all instances in S: The adversary either fails to go with the majority of the concepts in $\mathcal{V}$ on some $x \in S$ (which immediately halves $\mathcal{V}$) or goes with the majority on all $x \in S$. In the latter case, $\mathcal{V}$ is halved after all $|S|$ membership queries have been issued. •

Corollary 4.3 $HC_*^{EQ,MQ}(\mathcal{C},\mathcal{H}) = Hdim_*(\mathcal{C},\mathcal{H})$.

Theorem 4.4 $SHdim(\mathcal{V},\mathcal{H}) \leq HC^{EQ}(\mathcal{V},\mathcal{H}) \leq \lceil \ln(4) \cdot SHdim(\mathcal{V},\mathcal{H}) \rceil$.

Proof For sake of simplicity, set $q := \mathrm{HC}^{EQ}(\mathcal{V},\mathcal{H})$ and $d := \mathrm{SHdim}(\mathcal{V},\mathcal{H})$.

First, we show that $q \geq d$. The minimality of d implies that there exists a sample $C : X \to \{0,1,?\}$ without consistent explanation in $\mathcal{H}$ such that each subsample $S \sqsubseteq C$ of size at most $d-1$ is consistent with more than half of the concepts in $\mathcal{V}$. Thus, any halving algorithm issuing equivalence queries with hypotheses from $\mathcal{H}$ fails to be consistent with C and may obtain counterexamples taken from C. After up to $d-1$ queries, these counterexamples form a subsample $S \sqsubseteq C$ of size at most $d-1$. Hence, at least one additional query is needed for halving the initial version space $\mathcal{V}$.

In order to prove $q \leq \lceil \ln(4) \cdot d \rceil$, we describe an appropriate halving algorithm A. A keeps track of the current version space $\mathcal{W}$ (which is $\mathcal{V}$ initially). For $i = 0, 1$, let $M_{\mathcal{W}}^i$ be the set

$$\left\{ x \in X : \text{the fraction of concepts } C \in \mathcal{W} \text{ with } C(x) = 1 - i \text{ is less than } \frac{1}{2d} \right\}.$$

In other words, a very large fraction (at least $1 - 1/(2d)$) of the concepts in $\mathcal{V}$ votes for output label i on instances from $M_{\mathcal{W}}^i$. Let $M_{\mathcal{W}}$ be the sample assigning label $i \in \{0,1\}$ to all instances from $M_{\mathcal{W}}^i$ and label "?" to all remaining instances (those without a so clear majority). Let $S \sqsubseteq M_{\mathcal{W}}$ be an arbitrary but fixed subsample of size at most d. The definition of $M_{\mathcal{W}}^i$ implies (through some easy-to-check counting) that more than half of the concepts in $\mathcal{W}$ are consistent with S. The definition of the strong halving dimension implies that $M_{\mathcal{W}}$ has a consistent explanation, say $H_{\mathcal{W}}$, in $\mathcal{H}$.

The punchline of this discussion is: if A issues the the equivalence query with hypothesis $H_{\mathcal{W}}$ (for the current version space $\mathcal{W}$), then the next counterexample will shrink $\mathcal{W}$ by a factor $1 - 1/(2d)$ (or by a smaller factor). For the purpose of halving the initial version space $\mathcal{V}$, a sufficiently large number of equivalence queries is therefore obtained by solving

$$\left(1 - \frac{1}{2d}\right)^{q'} \cdot |\mathcal{V}| < e^{-q'/(2d)} \cdot |\mathcal{V}| \leq \frac{1}{2} \cdot |\mathcal{V}|$$

for q'. Clearly, $q = \lceil \ln(4) \cdot d \rceil$ is sufficiently large. •

Corollary 4.5 $SHdim_*(\mathcal{C}, \mathcal{H}) \le HC_*^{EQ}(\mathcal{C}, \mathcal{H}) \le \lceil \ln(4) \cdot SHdim_*(\mathcal{C}, \mathcal{H}) \rceil$.

5 An Application of the Halving Dimension

In this section, we show that the number equivalence queries needed to halve a given version spave $\mathcal{V}$ (roughly) equals the total number of equivalence and membership queries needed to halve $\mathcal{V}$ on the "hardest subdomain" K of X. Loosely speaking, there is always a subdomain K of X that leaves the problem of halving $\mathcal{V}$ by means of equivalence queries as as hard as before, but which renders membership queries useless. A similar result was shown in [2] for contraction (deterministic exact learning). However, this result left a gap of size $\theta(\log |\mathcal{C}|)$. The result proven here leaves only a (small) constant gap.

Let $K \subseteq X$. For each function $F : X \to \{0, 1\}$, let F_K denote the restriction of F to subdomain K. For each class $\mathcal{F}$ of functions from X to $\{0, 1\}$, we define $\mathcal{F}_K := \{F_K : F \in \mathcal{F}\}$. Then the following holds:

Lemma 5.1 $SHdim(\mathcal{V}, \mathcal{H}) = \max_{K \subseteq X} Hdim(\mathcal{V}_K, \mathcal{H}_K)$.

Proof Set $d := \mathrm{SHdim}(\mathcal{V}, \mathcal{H})$. Remember that d is the smallest number such that for all samples $S : X \to \{0, 1, ?\}$ the condition described in the second part of Definition 4.1 is satisfied. Let d_K be the corresponding number when we restrict ourselves to samples S with support K. It is evident that $d = \max_{K \subseteq X} d_K$ and $d_K = \mathrm{Hdim}(\mathcal{V}_K, \mathcal{H}_K)$, which completes the proof of the lemma. •

Combining Lemma 5.1, Corollary 4.3 and Corollary 4.5, we get

Corollary 5.2

$$\max_{K \subseteq X} HC^{EQ,MQ}(\mathcal{V}_K, \mathcal{H}_K) \le HC^{EQ}(\mathcal{V}, \mathcal{H}) \le \ln(4) \cdot \max_{K \subseteq X} HC^{EQ,MQ}(\mathcal{V}_K, \mathcal{H}_K).$$

Among the obvious open problems are the following-ones:

- The relation between halving and learning complexity that is proven in this paper leaves a gap. Can this gap be removed (at least for some concrete popular classes)?
- Can the (strong) halving dimension be determined for some popular concept and hypothesis classes?

References

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

[2] José L. Balcázar, Jorge Castro, David Guijarro, and Hans U. Simon. The consistency dimension and distribution-dependent learning from queries. *Theoretical Computer Science*, 288(2):197–215, 2002. An extended abstract of the paper appeared in the proceedings of ALT'99.

[3] Lisa Hellerstein, Krishnan Pillaipakkamnatt, Vijay Raghavan, and Dawn Wilkins. How many queries are needed to learn? *Journal of the Association on Computing Machinery*, 43(5):840–862, 1996.

[4] Wolfgang Maass. On-line with an oblivious environment and the power of randomization. In *Proceedings of the 4th Annual Workshop on Computational Learning Theory*, pages 167–175, 1991.