Lower Bounds for the Complexity of Learning Half-Spaces with Membership Queries

Valery N. Shevchenko and Nikolai Yu. Zolotykh

Nizhny Novgorod State University, Gagarin ave. 23,
Nizhny Novgorod 603600, Russia

Abstract. Exact learning of half-spaces over finite subsets of $\mathbb{R}^n$ from membership queries is considered. We describe the minimum set of labelled examples separating the target concept from all the other ones of the concept class under consideration. For a domain consisting of all integer points of some polytope we give non-trivial lower bounds on the complexity of exact identification of half-spaces. These bounds are near to known upper bounds.

1 Introduction

We consider the complexity of exact identification of half-spaces over the domain M that is an arbitrary finite subset of $\mathbb{R}^n$ (n is fixed). We are interested in the model of learning with membership queries.

The main result of this paper is Theorem 9 describing the structure of the teaching set T of a half-space c, i. e. a subset of M such that no other half-space agrees with c on the whole T.

The mentioned theorem is used to obtain the lower bound for the complexity of identification of half-spaces over the domain $\{0, 1, \ldots, k-1\}^n$. We show that $\mathrm{MEMB}(\mathrm{HS}_k^n) = \Omega(\log^{n-2} k)$. For $n \geq 3$ this significantly improves $\Omega(\log k)$ lower bound [10] on the considered quantity. The presented result can be compared with the following upper bound. From results of M. Yu. Moshkov in the test theory [11] it follows that

$$\mathrm{MEMB}(\mathrm{HS}_k^n) = O\left(\frac{\log^n k}{\log\log k}\right)$$

(see [8]). We remark that for any fixed n there is a learning algorithm that requires $O(\log^n k)$ membership queries and polynomial in $\log k$ running time. This algorithm was proposed in [20, 21, 8] .

When M is the set of all integer points of some polytope we give a lower bound for the complexity $\mathrm{MEMB}(\mathrm{HS}(M))$. We show that for any fixed n and $l > n$ and for any γ there is a polytope $P \subset \mathbb{R}^n$ described by a system of l linear inequalities with integer coefficients by absolute value not exceeding γ such that $\mathrm{MEMB}(\mathrm{HS}(P \cap \mathbb{R}^n)) = \Omega(l^{\lfloor n/2 \rfloor} \log^{n-1} \gamma)$. We remark that this bound is near to an upper bound obtained in the threshold function deciphering formalism: an algorithm that learns a half-space over $P \cap \mathbb{Z}^n$ in time bounded polynomially in

l and $\log \gamma$ using $O(l^{\lfloor n/2 \rfloor} \log^n \gamma)$ membership queries was proposed in [16] (n is fixed).

Some other related results see in Sect. 6.

2 Preliminaries

Let M is an arbitrary finite non–empty subset of $\mathbb{R}^n$. M is considered as an *instance space*. A *concept* over M is a subset of M. A *concept class* is some non-empty collection of concepts over M. The concept $c \subseteq M$ is called a *half-space* over M if there exist real numbers $a_0, a_1, \ldots, a_n$ such that

$$c = \left\{ x \in M \mid \sum_{j=1}^{n} x_j a_j \le a_0 \right\} \quad . \tag{1}$$

The inequality in (1) is called a *threshold inequality* for c. Denote by $\mathrm{HS}(M)$ the set of all half-spaces over M. Define $\mathrm{HS}_k^n = \mathrm{HS}(E_k^n)$ where $E_k = \{0, 1, \ldots, k-1\}$. Each half-space over M is a concept. The class $\mathrm{HS}(M)$ is a concept class.

We consider the model of *exact learning* [1, 10] with *membership queries*. The goal of the learner is to identify an unknown target concept c chosen from a known concept class C, making membership queries ("Is $x \in c$?" for some $x \in M$) and receiving yes/no answers. The complexity of a learning algorithm for C is the maximum number of queries it makes, over all possible target concepts $c \in C$. The complexity $\mathrm{MEMB}(C)$ of a concept class C is the minimum learning complexity, over all learning algorithms for this class. A set $T \subseteq M$ is said to be a *teaching set* for a concept $c \in C$ with respect to the class C if no other concept from C agrees with c on the whole T. If a teaching set is of minimum cardinality, over all teaching sets for a concept c, then we call it *minimum teaching set* for c. Denote by $\mathrm{TD}(c, C)$ the cardinality of a minimum teaching set for a concept c. $\mathrm{TD}(C)$ is maximum $\mathrm{TD}(c, C)$ over all concepts c in C. $\mathrm{TD}(C)$ is called *teaching dimension* for the class C. It is clear that $\mathrm{MEMB}(C) \ge \mathrm{TD}(C)$ (cf. [9]).

Let $\mathrm{Conv}\,(X)$ be the convex hull of $X \subseteq \mathbb{R}^n$; $\mathrm{Affdim}\,(X)$ is the affine dimension of X. For a concept $c \subseteq M$ denote by $N_0(c)$ (resp. $N_1(c)$) the set of vertices of $\mathrm{Conv}\,(c)$ (resp. $\mathrm{Conv}\,(M \setminus c)$). Denote $P_\nu(c) = \mathrm{Conv}\, N_\nu(c)$ ($\nu = 0, 1$).

3 Auxiliary Results

We first remark that a concept c over the domain M belongs to $\mathrm{HS}(M)$ if and only if $P_0(c) \cap P_1(c) = \emptyset$. Indeed, the necessity is evident and the sufficiency follows from the Separating Hyperplane Theorem (see [5]).

Associated with each half-space c over M is the cone $K(c)$ of separating functionals $a = (a_0, a_1, \ldots, a_n, a_{n+1})$ in an $(n+2)$ -dimensional vector space

[13, 15]; $K(c)$ is described by the conditions

$$\begin{cases} \sum\limits_{j=1}^{n} a_j x_j \le a_0 & \text{for each } x \in c \ , \\ \sum\limits_{j=1}^{n} a_j x_j \ge a_0 + a_{n+1} & \text{for each } x \in M \setminus c \ , \\ a_{n+1} \ge 0 \ . \end{cases} \tag{2}$$

Any solution $(a_0, \ldots, a_{n+1})$ of this system, with $a_{n+1} > 0$, defines a threshold inequality for c. The opposite is also true: the coefficients $(a_0, \ldots, a_n)$ of any threshold inequality of c satisfy the system (2) for some positive value of a_{n+1}.

For any $T_0 \subseteq c$, $T_1 \subseteq M \setminus c$ we consider the next subsystem of (2):

$$\begin{cases} \sum\limits_{j=1}^{n} a_j x_j \le a_0 & \text{for each } x \in T_0 \ , \\ \sum\limits_{j=1}^{n} a_j x_j \ge a_0 + a_{n+1} & \text{for each } x \in T_1 \ , \\ a_{n+1} \ge 0 \ . \end{cases} \tag{3}$$

Denote by $K(T_0, T_1)$ the cone consisting of its solutions. The set

$$K^*(T_0, T_1) = \left\{ \sum_{x \in T_0} \lambda_x \begin{pmatrix} 1 \\ -x \\ 0 \end{pmatrix} + \sum_{x \in T_1} \lambda_x \begin{pmatrix} -1 \\ x \\ -1 \end{pmatrix} + \nu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \mid \lambda_x \ge 0, \nu \ge 0 \right\}$$

is a cone, dual to $K(T_0, T_1)$. A cone is said to be *pointed* if it does not contain non-zero subspaces.

Lemma 1. *For any $T_0 \subseteq c$, $T_1 \subseteq M \setminus c$ the cone $K^*(T_0, T_1)$ is pointed.*

Proof. Since $0 \in K(T_0, T_1)$, for some non-negative ν and λ_x $(x \in T_0 \cup T_1)$ we have that $0 = \sum\limits_{x \in T_0} \lambda_x \cdot (1, -x, 0) + \sum\limits_{x \in T_1} \lambda_x \cdot (-1, x, -1) + \nu \cdot (0, 0, 1)$; consequently, $\sum\limits_{x \in T_0} \lambda_x = \sum\limits_{x \in T_1} \lambda_x = \nu$. If $\nu = 0$ then for any $x \in T_0 \cup T_1$ it holds that $\lambda_x = 0$, hence $K^*(T_0, T_1)$ is a pointed cone. If $\nu \neq 0$ then the point $y = \frac{1}{\nu} \sum\limits_{x \in T_0} \lambda_x x = \frac{1}{\nu} \sum\limits_{x \in T_1} \lambda_x x$, evidently, belongs to $P_0 \cap P_1$ that is impossible. □

Lemma 2. *For any $c \in \mathrm{HS}(M)$ the dimension of $K(c)$ is $n + 2$.*

Proof. It is known [5] that the cone K has the full dimension if and only if the dual cone K^* is pointed. Since $K(c) = K(c, M \setminus c))$, the assertion follows from Lemma 1. □

Lemma 3. *If $\mathrm{Affdim}\, M = n$ then for any $c \in \mathrm{HS}(M)$ the cone $K(c)$ is pointed.*

Proof. It is sufficient to verify that if $a = (a_0, a_1, \ldots, a_n, a_{n+1}) \in K(c)$ and $-a \in K(c)$ then $a = 0$. From the system (2) we get that in this case $a_{n+1} = 0$ and, consequently,

$$M \subseteq \left\{ x = (x_1, x_2, \ldots, x_n) \mid \sum_{j=1}^{n} a_j x_j = a_0 \right\} \ .$$

Since the dimension of M is n, all a_i $(i = 0, \ldots, n)$ are zeroes. $\square$

Now the following is a consequence of the theory of linear inequalities [5, 12].

Lemma 4. *If* $\operatorname{Affdim} M = n$ *then for every* $c \in \mathrm{HS}(M)$

1) the cone $K(c)$ *has a unique up to positive factors generating system (the system of extreme rays)*

$$\{\widetilde{b^{(i)}} = (b_0^{(i)}, b_1^{(i)}, \ldots, b_n^{(i)}, b_{n+1}^{(i)}),\ i = 1, \ldots, s\} \ ; \tag{4}$$

2) there are unique sets $T_0(c) \subseteq c$, $T_1(c) \subseteq M \setminus c$ *such that (2) is equivalent to the system*

$$\begin{cases} \sum\limits_{j=1}^{n} a_j x_j \le a_0 & \text{for each } (x_1, \ldots, x_n) \in T_0(c) \ , \\ \sum\limits_{j=1}^{n} a_j x_j \ge a_0 + a_{n+1} & \text{for each } (x_1, \ldots, x_n) \in T_1(c) \ , \\ a_{n+1} \ge 0 \end{cases} \tag{5}$$

and no subsystem of (5) is equivalent to the system (2);

3) for any $x = (x_1, \ldots, x_n) \in T_0(c)$ *there is a subset* $I \subseteq \{1, \ldots, s\}$ *such that* $|I| = n + 1$, *the system* $\{\widetilde{b^{(i)}}, i \in I\}$ *is linearly independent and*

$$\sum_{j=1}^{n} b_j^{(i)} x_j = b_0^{(i)} \ (i \in I), \ \sum_{i \in I} b_{n+1}^{(i)} > 0 \ ; \tag{6}$$

4) for any $x = (x_1, \ldots, x_n) \in T_1(c)$ *there is a subset* $I \subseteq \{1, \ldots, s\}$ *such that* $|I| = n + 1$, *the system* $\{\widetilde{b^{(i)}}, i \in I\}$ *is linearly independent and*

$$\sum_{j=1}^{n} b_j^{(i)} x_j = b_0^{(i)} + b_{n+1}^{(i)} \ (i \in I), \ \sum_{i \in I} b_{n+1}^{(i)} > 0 \ .$$

$\square$

There is the standard method to reduce the problem with $\operatorname{Affdim} M < n$ to the case of full dimension. Let $M \subseteq \mathbb{Q}^n$. Denote by Aff M the affine hull of M. Suppose that $\operatorname{Aff} M = \{x \in \mathbb{R}^n \mid Ax = b\}$ for some $A \in \mathbb{Z}^{m \times n}$. Let D be a Smith's normal diagonal matrix for A, the matrices P and Q are unimodular matrices such that $PAQ = D$. Without loss of generality we can take, $D = (I_m, 0)$ where I_m is an identity $m \times m$ matrix, 0 is a zero $n \times (n - m)$ matrix. Perform the change of variables $x = Qy$ mapping $\mathbb{Z}^n$ into $\mathbb{Z}^n$. We have that $PAx = Dy$, that is, Aff M is described by the conditions $y' = Pb$ where $y' = (y_1, \ldots, y_m)$. Thus, rewriting remaining conditions in variables $y'' = (y_{m+1}, \ldots, y_n)$ we get the problem in $\mathbb{R}^{n-m}$ with Affdim $M = n - m$. We remark that there exist P, Q such that the maximal by absolute value coefficient in the new problem does not exceed some polynomial in the maximal coefficient of the old problem (see, for example, [12]).

4 Caracterization of Teaching Sets of Half-Spaces

Theorem 5. *Let* $T_0 \subseteq c$, $T_1 \subseteq M \setminus c$. $T = T_0 \cup T_1$ *is a teaching set for a half-space c if and only if (3) is equivalent to (2).*

Proof. The sufficiency of the conditions is evident. We prove their necessity. Assume that there is the solution $b = (b_0, b_1, \ldots, b_n, b_{n+1})$ of (3) that does not belong to $K(c)$. By Lemma 2, we can suppose that $b_{n+1} > 0$. The threshold inequality $\sum_{j=1}^{n} b_j x_j \leq b_0$ defines some concept $g \in \mathrm{HS}(M)$. We have that $b \notin K(c)$, thus $g \neq c$. But g agrees with c on T. Hence T is not a teaching set. □

This theorem leads to

Corollary 6. *Let* $T_0 \subseteq c$, $T_1 \subseteq M \setminus c$, *then for any* $c \in \mathrm{HS}(M)$ *the set* $T = T_0 \cup T_1$ *is a minimum teaching set if and only if* $T_\nu = T_\nu(c)$ $(\nu = 0, 1)$. □

We note that the 2nd assertion of Lemma 4 is true for any $M \subseteq \mathbb{R}^n$, also when Affdim $M < n$. By Corollary 6, we now get

Corollary 7. *For any* $c \in \mathrm{HS}(M)$ *there is a unique minimum teaching set. It is contained in every teaching set of c.* □

Denote by $T(c) = T_0(c) \bigcup T_1(c)$ the minimum teaching set for c.

Corollary 8. *(Cf. [14, 7]) For any* $c \in \mathrm{HS}(M)$ *it holds that* $T(c) \subseteq N_0(c) \cup N_1(c)$.

Proof. It is obvious that for $T_\nu = N_\nu(c)$ the system (3) is equivalent to the system (2). The assertion of the corollary follows now from Theorem 5. □

Let Affdim $M = n$ and $c \in \mathrm{HS}(M)$. Without loss of generality we can assume that in (4) it holds that $b_{n+1}^{(i)} > 0$ for any $i = 1, \ldots, \mu$ and $b_{n+1}^{(i)} = 0$ for any $i = \mu + 1, \ldots, s$. Let $a = (a_1, \ldots a_n)$,

$$M_0(c, a) = \left\{ (y_1, \ldots, y_n) \in M \mid \sum_{j=1}^{n} a_j y_j = \max_{x \in c} \sum_{j=1}^{n} a_j x_j \right\} ,$$

$$M_1(c, a) = \left\{ (y_1, \ldots, y_n) \in M \mid \sum_{j=1}^{n} a_j y_j = \min_{x \in M \setminus c} \sum_{j=1}^{n} a_j x_j \right\} .$$

Denote by $N_\nu(c, a)$ the set of vertices of the convex hull of $M_\nu(c, a)$.

Theorem 9. *If* Affdim $M = n$ *then for any* $c \in \mathrm{HS}(M)$ *it holds that*

$$T(c) = \bigcup_{i=1}^{\mu} \left(N_0(c, \widetilde{b^{(i)}}) \cup N_1(c, \widetilde{b^{(i)}}) \right) = \bigcup_{a} \left(N_0(c, a) \cup N_1(c, a) \right) \quad ,$$

in the right-hand side the union is over all $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ *such that the inequality*

$$\sum_{j=1}^{n} a_j x_j \le \max_{x \in c} \sum_{j=1}^{n} a_j x_j$$

is a threshold inequality for c.

Proof. First we prove the inclusion $T(c) \subseteq \bigcup\limits_{i=1}^{\mu} \left(N_0(c, \widetilde{b^{(i)}}) \cup N_1(c, \widetilde{b^{(i)}}) \right)$. Let $y = (y_1, \ldots y_n) \in T_0(c)$. By the 3rd assertion of Lemma 4, there is $i \in \{1, \ldots, \mu\}$ such that $\sum\limits_{j=1}^{n} b_j^{(i)} y_j = b_0^{(i)}$. Since $b_{n+1}^{(i)} > 0$, the coefficients $b_j^{(i)}$ $(j = 0, 1, \ldots, n)$ are the coefficients of a threshold inequality for c and $\max\limits_{x \in c} \sum\limits_{j=1}^{n} x_j b_j^{(i)} = b_0^{(i)}$. It follows from this that $y \in M_0(c, \widetilde{b^{(i)}})$. Assume that $y \notin N_0(c, \widetilde{b^{(i)}})$, i. e. $y = \sum\limits_{q=1}^{p} \alpha_q y^{(q)}$ for some $p > 1$, $\alpha_q > 0$, $\sum\limits_{q=1}^{p} \alpha_q = 1$, $y \neq y^{(q)} \in M_0(c, \widetilde{b^{(i)}})$ $(q = 1, \ldots, p)$. Then $y \notin N_0(c)$ and, by Corollary 8, $y \notin T_0(c)$. This contradiction shows that $y \in N_0(c, \widetilde{b^{(i)}})$. The case $y \in T_1(c)$ is proved similarly by the 4th assertion of Lemma 4.

We now prove that $\bigcup\limits_{a} (N_0(c, a) \cup N_1(c, a)) \subseteq T(c)$. Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $a_0 = \max\limits_{x \in c} \sum\limits_{j=1}^{n} a_j x_j$; $\sum\limits_{j=1}^{n} a_j x_j \le a_0$ is a threshold inequality for c. For any point $z \in N_0(c, a)$ we consider a concept $g = c \setminus \{z\}$. Let us prove that $g \in \mathrm{HS}(M)$. Assume the contrary, then $P_0(g) \bigcap P_1(g) \neq \emptyset$. This means that there are points $x^{(1)}, \ldots, x^{(p)}$ in g, points $y^{(0)}, \ldots, y^{(q)}$ in $M \setminus g$, and positive numbers $\alpha_1, \ldots \alpha_p$, $\beta_0, \ldots \beta_q$ such that

$$x = (x_1, \ldots, x_n) = \sum_{r=1}^{p} \alpha_r x^{(r)} = \sum_{t=0}^{q} \beta_t y^{(t)} \quad , \tag{7}$$

$\sum\limits_{r=1}^{p} \alpha_r = 1$, $\sum\limits_{t=0}^{q} \beta_t = 1$ where $x \in P_0(g) \cap P_1(g)$. It is clear that among $y^{(0)}, \ldots, y^{(q)}$ there is a point z, since otherwise we obtain that $P_0(c) \bigcap P_1(c) \neq \emptyset$, that is impossible, because it holds that $c \in \mathrm{HS}(M)$. Let $z = y^{(0)}$. We have that $\sum\limits_{j=1}^{n} a_j x_j = \sum\limits_{r=1}^{p} \alpha_r \sum\limits_{j=1}^{n} a_j x_j^{(r)} = \sum\limits_{t=1}^{q} \beta_t \sum\limits_{j=1}^{n} a_j y_j^{(t)} + \beta_0 \sum\limits_{j=1}^{n} a_j z_j$. In the last formula the central part does not exceed a_0; in the right-hand side the first addend is greater than a_0, and the second one is equal to a_0. For the equality it is necessary that $q = 0$ and $\sum_{j=1}^{n} a_j x_j^{(r)} = a_0$ $(r = 1, \ldots, p)$. Thus, $\beta_0 = 1$, $z = x$. From (7) we now obtain that $z \notin N_0(c, a)$, that contradicts the condition. Hence $g \in \mathrm{HS}(M)$. Since c and g differ only at one point, we have that $z \in T(c)$.

Suppose now that $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $a_0 = \min_{x \in M \setminus c} \sum_{j=1}^{n} a_j x_j$. The inequality $\sum_{j=1}^{n} a_j x_j \geq a_0$ is true for any point in $M \setminus c$ and it is false for any point in c. For each $z \in N_1(c, a)$ we define a concept $g = c \cup \{z\}$. The further proof is the same one described above.

It is obvious that $\bigcup_{i=1}^{\mu} \left(N_0(c, \widetilde{b^{(i)}}) \cup N_1(c, \widetilde{b^{(i)}}) \right) \subseteq \bigcup_{a} (N_0(c, a) \cup N_1(c, a))$. The last inclusion finishes the proof of the theorem. □

The example $x \in M$ is called *essential* for a concept $c \in \mathrm{HS}\,(M)$ if there is $g \in \mathrm{HS}\,(M)$ such that c and g agree on $M \setminus \{x\}$ and don't agree at the point x. From the last part of Theorem 9 it follows that $T(c)$ is exactly the set of essential examples for c. For the case of Boolean domain E_2^n this is a well-known result (see [2] and related papers referenced in [2]).

As an example of Theorem 9, consider the concept $c \in \mathrm{HS}(E_9^3)$ defined by the threshold inequality $20x_1 + 28x_2 + 35x_3 \leq 140$. Rewrite the system (5) as $Qa \geq 0$ where $a = (a_0, \ldots, a_{n+1})^{\mathrm{T}}$ is a column of variables and Q is a matrix formed from the coordinates of the points of $T(c)$. Let B be a matrix formed from the entries of the vectors $\widetilde{b^{(i)}}$, $S = QB$, I is an identity matrix. The matrix

$$\left(\begin{array}{c|c} E & B \\ \hline Q & S \end{array} \right)$$

is represented in Table 4. We have that $\mu = 3$,

$$N_0(c, \widetilde{b^{(1)}}) = \{p^{(1)}, p^{(3)}\}, \quad N_1(c, \widetilde{b^{(1)}}) = \{q^{(1)}, q^{(2)}\} \ ,$$

$$N_0(c, \widetilde{b^{(2)}}) = \{p^{(1)}, p^{(2)}\}, \quad N_1(c, \widetilde{b^{(2)}}) = \{q^{(1)}, q^{(3)}\} \ ,$$

$$N_0(c, \widetilde{b^{(3)}}) = \{p^{(1)}, p^{(2)}, p^{(3)}\}, \quad N_3(c, \widetilde{b^{(1)}}) = \{q^{(1)}\}$$

where $p^{(1)} = (7, 0, 0)$, $p^{(2)} = (0, 5, 0)$, $p^{(3)} = (0, 0, 4)$, $q^{(1)} = (4, 1, 1)$, $q^{(2)} = (3, 3, 0)$, $q^{(3)} = (2, 0, 3)$, $\widetilde{b^{(1)}} = (56, 8, 11, 14, 1)$, $\widetilde{b^{(2)}} = (70, 10, 14, 17, 1)$, $\widetilde{b^{(3)}} = (140, 20, 28, 35, 140)$. By Theorem 9, $T_\nu(c) = \bigcup_{i=1}^{3} N_\nu(c, \widetilde{b^{(i)}})$ $(\nu = 0, 1)$. For the considered example in the union it suffices to retain solely 2 members. Indeed, $T_0(c) = N_0(c, \widetilde{b^{(3)}}) = N_0(c, \widetilde{b^{(1)}}) \cup N_0(c, \widetilde{b^{(2)}})$, $T_1(c) = N_1(c, \widetilde{b^{(1)}}) \cup N_1(c, \widetilde{b^{(2)}})$.

Table 1. Example of Theorem 9

1	0	0	0	0	56	70	140	140	105	84	80	50	36	21
0	1	0	0	0	8	10	20	20	15	12	11	7	5	3
0	0	1	0	0	11	14	28	28	21	16	16	10	7	4
0	0	0	1	0	14	17	35	35	25	21	20	12	9	5
0	0	0	0	1	1	1	3	0	0	0	0	0	0	0
1	-7	0	0	0	0	0	0	0	0	0	3	1	1	0
1	0	-5	0	0	1	0	0	0	0	4	0	0	1	1
1	0	0	-4	0	0	2	0	0	5	0	0	2	0	1
-1	4	1	1	-1	0	0	0	3	1	1	0	0	0	0
-1	3	3	0	-1	0	1	1	4	3	0	1	1	0	0
-1	2	0	3	-1	1	0	2	5	0	3	2	0	1	0
0	0	0	0	1	1	1	3	0	0	0	0	0	0	0

5 Bounds for the Teaching Dimension of Half-Spaces

Denote by N the set of vertices of the polytope $\mathrm{Conv}\, M$.

Lemma 10. *If $c = M$ or $c = \emptyset$ then it holds that $T(c) = N$.*

Proof. Assume that for $c = M$ there is a point $x \in N \setminus T(c)$. Consider the concept $g = M \setminus \{x\}$. Since $x \in N$, it is clear that $g \in \mathrm{HS}(M)$ and, consequently, $x \in T(c)$. We have proved that $N \subseteq T(c)$. The opposite inclusion follows from Corollary 8. For $c = \emptyset$ the lemma can be proved by analogy. □

From Lemma 10 it follows that $\mathrm{TD}(\mathrm{HS}_k^n) \geq 2^n$. Indeed, assume that $c = E_k^n$. By Lemma 10 we have that $\mathrm{TD}(c) = 2^n$, hence $\mathrm{TD}(\mathrm{HS}_k^n) \geq 2^n$. Thus, no polynomial in n algorithm for learning half-spaces over E_k^n from membership queries exists. This was originally proved in [14].

Let P be a polytope in $\mathbb{R}^n$ that can be described as an integer system of l linear inequalities with integer coefficients whose absolute values do not exceed γ. Denote by $\mathcal{P}(n, l, \gamma)$ the class of all such polytopes. For the class $\mathrm{HS}(M)$ with $M = P \cap \mathbb{Z}^n$ and $P \in \mathcal{P}(n, l, \gamma)$ we have

Theorem 11. *For every natural $n \geq 2$ and $l > n$ there is γ_0 such that for every $\gamma \geq \gamma_0$ there exists a polytope $P \in \mathcal{P}(n, l, \gamma)$ such that*

$$\mathrm{MEMB}(\mathrm{HS}(M)) \geq \mathrm{TD}(\mathrm{HS}(M)) \geq D_n l^{\lfloor n/2 \rfloor} \log^{n-1} \gamma$$

where $M = P \cap \mathbb{Z}^n$ and D_n is some positive quantity depending only on n.

Proof. It was proved in [6] (cf. [3]) that for any fixed $n \geq 2$ and $l > n$, for any sufficiently large γ there exists a polytope $P \in \mathcal{P}(n, l, \gamma)$ such that the number of vertices of $\mathrm{Conv}\,(P \bigcap \mathbb{Z}^n)$ is not less than $D_n l^{\lfloor n/2 \rfloor} \log^{n-1} \gamma$. The assertion to be proved follows now from Lemma 10. □

Return to the class HS_k^n. Denote by $N(a_0, a_1, \ldots, a_n)$ the set of all vertices of a convex hull of solutions of the following system:

$$\begin{cases} \sum_{j=1}^{n} a_j x_j = a_0 \ ; \\ x_j \geq 0; \ x_j \in \mathbb{Z} \ (j = 1, \ldots, n) \ . \end{cases}$$

In [18] S. I. Veselov got a lower bound for the mean quantity of $|N(a_0, a_1, \ldots, a_n)|$ (see Sect. 3.5 of [15]). This leads to

Lemma 12. *For every $n \geq 2$, $k \geq 2$ there are positive numbers $a_0, a_1, \ldots, a_n$ such that $a_i \leq k - 1$ $(i = 0, 1, \ldots, n)$ and*

$$|N(a_0, a_1, \ldots, a_n)| \geq C_n \log^{n-2} k$$

where C_n is some positive quantity depending only on n. □

Theorem 13. *For every $n \geq 2$ and $k \geq 2$*

$$C_n \log^{n-2} k \leq \mathrm{TD}(\mathrm{HS}_k^n) \leq C'_n \log^{n-1} k$$

where C_n and C'_n are some quantities depending only on n.

Proof. The lower bound was announced (without a proof) in [17]. To obtain it we construct a concept c in the following manner. Consider $a_0, a_1, \ldots, a_n$ in the assertion of Lemma 12 as the coefficients of a threshold inequality of c. Since $1 \leq a_i \leq k - 1$, we have that $N(a_0, \ldots, a_n) \subseteq E_k^n$. From Theorem 9 it follows that $T(c) \supseteq N(a_0, \ldots, a_n)$, hence, $\mathrm{TD}(c, \mathrm{HS}_k^n) \geq C_n \log^{n-2} k$, consequently, $\mathrm{TD}(\mathrm{HS}_k^n) \geq C_n \log^{n-2} k$.

The upper bound was proved by T. Hegedüs [7] on the base of [13]. It is clear that for $T_\nu = N_\nu(c)$ the system (2) is equivalent to the system (3), hence $T(c) \subseteq N_0(c) \bigcup N_1(c)$; it is known [7] that $|N_0(c)| + |N_1(c)| \leq C'_n \log^{n-1} k$ where C'_n is some quantity depending only on n. Thus for any concept $c \in \mathrm{HS}_n^k$ the inequality $\mathrm{TD}(c) \leq C'_n \log^{n-1} k$ holds. □

The lower bound in Theorem 13 gives us that $\mathrm{MEMB}(\mathrm{HS}_k^n) \geq C_n \log^{n-2} k$.

6 Related Results and Open Problems

In proving the lower bound for the teaching dimension of half-spaces over E_k^n we used the fact that the quantity μ in Theorem 9 is at least 1. An open problem remains: it would be helpful to estimate from *above* the quantity μ (we remark that for $n \geq 3$ there are examples with $\mu = 2, 3$). In this way one could apparently decrease the upper bound on $\mathrm{TD}(\mathrm{HS}_k^n)$. For instance, it is known from [17] that $\mathrm{TD}(\mathrm{HS}_k^2) = 4$. This result is of considerable interest because (as it was shown in [4, 19]) $\mathrm{MEMB}(\mathrm{HS}_k^2) = \Theta(\log k)$.

References

1. Angluin, D.: Queries and concept learning. Machine Learning (2) (1988) 319–342
2. Anthony, M., Brightwell, G., Shawe-Taylor, J.: On specifying Boolean functions by labelled examples. Discrete Applied Mathematics **61** (1) (1995) 1–25
3. Bárány, I., Howe, R., Lovász, L.: On integer points in polyhedra: a lower bound. Combinatorica (12) (1992) 135–142
4. Bultman, W. J., Maass, W.: Fast identification of geometric objects with membership queries. Information and Computation **118** (1) (1995) 48–64
5. Chernikov, S. N.: Linear Inequalities. "Nauka" Moscow (1968). German transl.: VEB Deutscher Verlag Wiss. Berlin (1971)
6. Chirkov, A. Yu.: On lower bound of the number of vertices of a convex hull of integer and partially integer points of a polyhedron. Proceedings of the First International Conference "Mathematical Algorithms". NNSU Publishers Nizhny Novgorod (1995) 128–134 (Russian)
7. Hegedüs, T.: Geometrical concept learning and convex polytopes. Proceedings of the 7th Annual ACM Conference on Computational Learning Theory (COLT'94). ACM Press New York (1994) 228–236
8. Hegedüs, T.: Generalized teaching dimensions and the query complexity of learning. Proceedings of the 8th Annual ACM Conference on Computational Learning Theory (COLT'95). ACM Press New York (1995) 108–117
9. Korobkov, V. K.: On monotone functions of logic algebra. Cybernetics Problems. "Nauka" Moscow **13** (1965) 5–28 (Russian)
10. Maass, W, Turán, Gy.: Lower bound methods and separation results for on-line learning models. Machine Learning (9) (1992) 107–145
11. Moshkov, M. Yu.: Conditional tests. Cybernetics Problems. "Nauka" Moscow **40** (1983) 131–170 (Russian)
12. Schrijver, A.: Theory of Linear and Integer Programming. Wiley–Interscience New York (1986)
13. Shevchenko, V. N.: On some functions of many–valued logic connected with integer programming. Methods of Discrete Analysis in the Theory of Graphs and Circuits. Novosibirsk **42** (1985) 99–102 (Russian)
14. Shevchenko, V. N.: Deciphering of a threshold function of many–valued logic. Combinatorial–Algebraic Methods in Applied Mathematics. Gorky (1987) 155–163 (Russian)
15. Shevchenko, V. N.: Qualitative Topics in Integer Linear Programming. "Fizmatlit" Moscow (1995). English transl.: AMS Providence Rhode Island (1997)
16. Shevchenko, V. N., Zolotykh, N. Yu.: Decoding of threshold functions defined on the integer points of a polytope. Pattern Recognition and Image Analysis. MAIK/Interperiodica Publishing Moscow **7** (2) (1997) 235–240
17. Shevchenko, V. N., Zolotykh, N. Yu.: On complexity of deciphering threshold functions of k-valued logic. Russian Math. Dokl. (Doklady Rossiiskoi Akademii Nauk) (to appear)
18. Veselov, S. I.: A lower bound for the mean number of irreducible and extreme points in two discrete programming problems. Manuscript No. 619–84, deposited at VINITI Moscow (1984) (Russian).
19. Zolotykh, N. Yu.: An algorithm of deciphering a threshold function of k-valued logic in the plane with the number of calls to the oracle $O(\log k)$. Proceedings of the First International Conference "Mathematical Algorithms". NNSU Publishers Nizhny Novgorod (1995) 21–26 (Russian)

20. Zolotykh, N. Yu., Shevchenko, V. N.: On complexity of deciphering threshold functions. Discrete Analysis and Operations Research. Novosibirsk **2** (1) (1995) 72–73 (Russian)
21. Zolotykh, N. Yu., Shevchenko, V. N.: Deciphering threshold functions of k-valued logic. Discrete Analysis and Operations Research. Novosibirsk **2** (3) (1995) 18–23. English transl.: Korshunov, A. D. (ed.): Operations Research and Discrete Analysis. Kluwer Ac. Publ. Netherlands (1997) 321–326

This article was processed using the LaTeX macro package with LLNCS style