Machine Learning, 14, 251–269 (1994)
© 1994 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Algorithms and Lower Bounds for On-Line Learning of Geometrical Concepts

WOLFGANG MAASS (maass@igi.tu-graz.ac.at)
Institute for Theoretical Computer Science, Technische Universität Graz, Klosterwiesgasse 32/2, A-8010 Graz, Austria

GYÖRGY TURÁN (U11557@UICVM.BITNET)
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60680; and Automata Theory Research Group of the Hungarian Academy of Sciences, Szeged, Hungary

Editor: David Haussler

Abstract. The complexity of on-line learning is investigated for the basic classes of geometrical objects over a discrete ("digitized") domain. In particular, upper and lower bounds are derived for the complexity of learning algorithms for axis-parallel rectangles, rectangles in general position, balls, halfspaces, intersections of halfspaces, and semi-algebraic sets. The learning model considered is the standard model for on-line learning from counterexamples.

Keywords. on-line learning, computational learning theory, geometrical objects

1. Introduction

The goal of this article is to investigate the complexity of on-line learning for the basic classes of geometrical objects over a discrete domain. The learning model that we consider is the most common one for on-line learning (introduced by Angluin, 1988). It may be viewed as a machine-independent version of the classical paradigma for learning from mistakes on perceptrons (Rosenblatt, 1962; Nilsson, 1965; Minsky & Papert, 1988) and neural networks (Nilsson, 1965; Rumelhart & McClelland, 1986; Lippmann, 1987).

A learning process in the learning model that we consider is a dialog between the "learner" and the "environment." The learner proposes "hypotheses" H from a fixed "concept class" $C \subseteq 2^X$ over a finite domain X. The goal of the learner is to "learn" an unknown "target concept" $C_T \in C$ that has been fixed by the environment. Whenever the learner proposes some hypothesis $H \in C$ with $H \neq C_T$, the environment responds with some "counterexample" $x \in H \triangle C_T := (C_T - H) \cup (H - C_T)$. The counterexample x is called a "positive counterexample" if $x \in C_T - H$, and x is called a "negative counterexample" if $x \in H - C_T$. A *learning algorithm for* C is any algorithm A that produces new hypotheses

$$H_{i+1}^A := A(x_1, \ldots, x_i; H_1^A, \ldots, H_i^A)$$

in dependence of counterexamples $x_j \in H_j^A \triangle C_T$ for the preceding hypotheses H_j^A. One also refers to these hypotheses as "equivalence queries" (Angluin, 1988).

The "learning complexity" LC(A) of such a learning algorithm A is defined by

$$\mathrm{LC}(A) := \max\{i \in \mathrm{N} \mid \text{there is some } C_T \in C \text{ and some choice of counterexamples } x_j \in H_j^A \triangle C_T \text{ for } j = 1, \ldots, i-1 \text{ such that } H_i^A \neq C_T\}.$$

The "learning complexity" LC(C) of a concept class C is defined by

$$\mathrm{LC}(C) := \min\{\mathrm{LC}(A) \mid A \text{ is a learning algorithm for } C\}.$$

Thus, analogous to the analysis of algorithms for computational problems, one carries out a worst-case analysis for each learning algorithm A. The learning complexity LC(C) of the concept class C is then defined as the learning complexity of the best learning algorithm A for C.

One says that a hypothesis H is a *consistent* with a positive (negative) counterexample $x \in X$ if $x \in H$ ($x \notin H$).

The concept classes C that are considered in this article are classes of "digitized" versions of basic geometric objects (similar to Minsky & Papert, 1988). For any fixed finite dimension d, we fix as domain the set $X_n^d := \{0, \ldots, n-1\}^d$ (one may view X_n^d as the set of "pixels" of a digital image representation device). Over this domain we consider the following concept classes:

$$\begin{aligned}
\mathrm{BOX}_n^d := {} & \{C \subseteq X_n^d \mid \text{there is a } d\text{-dimensional axis-parallel rectangle } R \subseteq \mathbf{R}^d \text{ with } R \cap X_n^d = C\} \\
= {} & \{\textstyle\times_{k=1}^{d} \{i_k, \ldots, j_k\} \mid 0 \leq i_k \leq j_k \leq n-1 \text{ for } k = 1, \ldots, d\} \cup \{\emptyset\}, \\
\text{GP-BOX}_n^d := {} & \{C \subseteq X_n^d \mid \text{there is a } d\text{-dimensional rectangle } R \subseteq \mathbf{R}^d \text{ with } R \cap X_n^d = C \ (R \text{ need not be axis parallel})\}, \\
\mathrm{HALFSPACE}_n^d := {} & \{C \subseteq X_n^d \mid \text{there is a halfspace } F \subseteq \mathbf{R}^d \text{ with } F \cap X_n^d = C\}, \\
\text{2-HALFSPACE}_n^d := {} & \{C \cap C' \mid C, C' \in \mathrm{HALFSPACE}_n^d\}, \\
\mathrm{BALL}_n^d := {} & \{C \subseteq X_n^d \mid \text{there is a ball } B \subseteq \mathbf{R}^d \text{ with } B \cap X_n^d = C\},
\end{aligned}$$

and k-SEMI-ALGEBRAIC-SETS$_n^d$ (defined in section 4).

As an aside, we would like to mention that concepts from the concept class BOX$_n^d$ also occur in contexts other than geometry. As an example, we would like to point to the hypothetical situation where one wants to learn from counterexamples the target concept "average-built person," which may be defined as the set of all persons whose weight lies in a certain interval $[w_1, w_2]$ and whose size lies in a certain interval $[s_1, s_2]$ (with unknown parameters w_1, w_2, s_1, s_2).

We show in this article that for any fixed dimension $d \geq 2$ one has LC(BOX_n^d) $= \Theta$ ($\log n$) (section 2), LC(HALFSPACE_n^d) $= \Theta(\log n)$ (section 4), LC(k-SEMI-ALGEBRAIC-SETS_n^d) $= \Theta(\log n)$ for any fixed $k \in$ N (section 4), and LC(BALL_n^d) $= \Theta(\log n)$ (section 5). All these upper bounds are realized by computationally feasible algorithms.

On the other hand, we show that for any $d \geq 2$, GP-BOX_n^d and 2-HALFSPACE_n^d require exponentially more learning steps than the classes mentioned above: one has for $d = 2$ LC(GP-BOX_n^d) $= \Omega(n)$ (section 3) and LC(2-HALFSPACE_n^d) $= \Omega(n)$ (section 4). Finally, in section 6 we present a list of open problems.

It turns out that for those concept classes C for which we prove LC(C) $= \Theta(\log n)$, the derived bound $\Theta(\log n)$ is very robust with regard to changes of the model for on-line learning. We prove the upper bound $O(\log n)$ in the rather weak model where the learner may only propose hypotheses from C. However, the lower bound $\Omega(\log n)$ also holds for the strongest model for on-line learning where the learner may propose arbitrary subsets $H \subseteq X$ as hypotheses, and where he can also ask membership queries. In fact, one even has LC-PARTICAL(C) $= \Omega(\log(\text{chain}(C)) = \Omega(\log n)$ for those classes C (see Maass & Turán, 1989, 1992) (in the model LC-PARTIAL, the learner may ask arbitrary hypotheses with "don't cares" for arbitrary elements of the domain; chain (C) is the length of the longest chain in C under inclusion).

The parameter $\log n$ denotes (up to constant factors) the size of an instant, i.e., the number of bits needed to specify an arbitrary point in the underlying domain X_n^d. Thus a concept class C may be viewed as "polynomially on-line learnable" only if LC(C) is polynomial in $\log n$. Hence, one may interpret the results of the present article as saying that rectangles in general position and intersections of halfspaces are not "polynomially learnable" in the on-line learning model considered, whereas all the other classes of geometrical objects mentioned above are polynomially (in fact: linearly) on-line learnable.

For any fixed dimension d, the concept classes C that are considered in this article have a constant VC-dimension. Therefore *any* computationally feasible algorithm that assigns to an arbitrary set of positive and negative examples (for a target concept $C_T \in C$) some concept $C \in C$ that is consistent with these examples provides for these concept classes C a polynomial learning algorithm in the pac-learning model (see Blumer, Ehrenfeucht, Haussler, & Warmuth, 1989; Haussler, Kearns, Littlestone, & Warmuth, 1991). It also provides a satisfactory prediction strategy in the associated probabilistic prediction model, where one assumes that the probability distribution over instances is time invariant (Haussler, Littlestone, Warmuth, 1987, 1988; Haussler, Kearns, Littlestone, & Warmuth, 1991). In contrast, it is easy to see that the common algorithms for assigning to a sequence of examples some consistent concept $C \in C$ (e.g., the OCCAM-algorithm that assigns the smallest consistent $C \in C$ in the case of BOX_n^d; see section 2) are not sufficient to provide a polynomial learning algorithm for on-line learning in a non-stochastic setting. Instead, an efficient *on-line* learning algorithm for the considered concept classes C has to issue hypotheses H that provide a more sophisticated interpolation between the available positive and negative (counter-) examples (in this respect, the hypotheses of an efficient *on-line* learning algorithm resemble more closely the kinds of hypotheses that a human learner might propose).

The learning algorithms that we present in this article are computationally feasible and consistent (i.e., they only issue hypotheses that are consistent with all preceding counter-examples). Hence they may also be used as efficient learning algorithms in the pac-learning model, or in the associated probabilistic prediction model (Haussler, Littlestone, & Warmuth, 1987, 1988). In fact, these learning algorithms appear to be preferable to other efficient pac-learning algorithms for the probabilistic prediction model. They have in addition to their favorable average-case behavior an optimal worst-case behavior. Hence they are more robust than other pac-learning algorithms insofar as they also provide good error bounds if the relatively strong assumptions of this probabilistic model are not met (in particular, if the underlying probability distribution of examples changes over the considered time period). Furthermore their (absolute) error bound of $O(\log n)$ is lower than the upper bound on the number of errors among m trials that can be derived for an arbitrary consistent pac-learning algorithm (see Haussler, Littlestone, & Warmuth, 1987, 1988), provided that m is not too small.

Some of the results in this article have previously been announced in the extended abstract of Maass and Turán (1989).

2. Learning of axis-parallel rectangles

In the first theorem, we consider the concept class

$$\mathrm{BOX}_n^d := \{ \mathop{\times}_{k=1}^{d} \{i_k, \ldots, j_k\}\} \mid 0 \leq i_k \leq j_k \leq n - 1$$
$$\text{for } k = 1, \ldots, d\} \cup \{\emptyset\},$$

which consists of all rectangular axis-parallel "boxes" that are contained in the discrete d-dimensional space $X_n^d := \{0, \ldots, n - 1\}^d$. It is obvious that $\mathrm{BOX}_n^d = \{C \subseteq X_n^d \mid$ there is a d-dimensional axis-parallel rectangle $R \subseteq \mathbf{R}^d$ with $R \cap X_n^d = C\}$.

Remark. Perhaps the simplest on-line learning algorithm for BOX_n^d is the "Occam-algorithm," which always chooses as next hypothesis the smallest $C \in \mathrm{BOX}_n^d$ that is consistent with all preceding counterexamples (see Blumer, Ehrenfeucht, Haussler, & Warmuth, 1989). It is easy to see that this algorithm needs in the worst case $\Omega(n)$ learning steps for BOX_n^d (for any dimension $d \geq 1$): choose as (positive) counterexample always a point that is just outside of the proposed hypothesis.

Theorem 1. Consider any fixed dimension $d \in \mathbf{N} - \{0\}$. Then $\mathrm{LC}(\mathrm{BOX}_n^d) = \Theta(\log n)$. Furthermore, there exists a learning algorithm A for BOX_n^d with $\mathrm{LC}(A) = O(\log n)$ that uses altogether at most $O(\mathrm{poly}(\log n))$ computation steps.

Proof. In order to design a learning algorithm for BOX_n^d that learns substantially faster than the naive algorithm (which always outputs the minimal consistent hypothesis), one has to generate hypotheses that interpolate between the minimal consistent hypothesis and some maximal consistent hypothesis.

In the case $d = 1$, there exists always a *unique* maximal consistent hypothesis (provided that some point in C_T is already known). Therefore it is trivial to construct for $d = 1$ a learning algorithm A with $\mathrm{LC}(A) = O(\log n)$: the next hypothesis of A always interpolates halfway between the minimal and the maximal consistent hypothesis ("binary search").

This method cannot be generalized to the case $d > 1$, since for $d > 1$ there is in general no unique maximal box that is consistent with the previously received counterexamples. For $d = 2$, some maximal consistent box may run *to the right* of some negative counterexample x, while another one avoids x by running *below* x to the left. This ambiguity corresponds to conflicting "theories" *why* x is not in the target box (or more precisely, which of the defining conditions for points in the target box are not met by x). For BOX_n^d, as well as for most other concrete concept classes that are discussed below, the interesting point in the design of an efficient learning algorithm lies in the construction of a next hypothesis H that guarantees substantial progress (from any counterexample to H), no matter which of the conflicting "theories" about the explanation of the previously received counterexamples are true.

Technically, this amounts to giving the right definition of "progress" for learning in the considered concept class.

For BOX_n^d, it is useful to measure the learning progress in terms of the number of points in X_n^d that could be a corner-point of the target box (on the basis of all counterexamples received so far).

In the following, we write $[i, j]$ for the set $\{m \in \mathbf{N} \mid i \leq m \leq j\}$ (we assume that $0 \in \mathbf{N}$). For an arbitrary box

$$C = \bigtimes_{k=1}^{d} [a_k, b_k] \in \mathrm{BOX}_n^d$$

and an arbitrary binary string $v = \langle v, \ldots, v_d \rangle \in \{0, 1\}^d$, we write C_v for the "v-corner" $\langle h_1, \ldots, h_d \rangle \in \{0, \ldots, n - 1\}^d$ of C, which is defined by

$$h_k = \begin{cases} a_k, \text{ if } v_k = 0 \\ b_k, \text{ if } v_k = 1. \end{cases}$$

At any point during a learning process, we write S_v for the set of all points in X_n^d that could still be the v-corner of the target concept, i.e.,

$$S_v = \{C_v \mid C \in \mathrm{BOX}_n^d \text{ is consistent with all counterexamples received so far}\}.$$

The next hypothesis $H \in \mathrm{BOX}_n^d$ of the learning algorithm is defined in such a way that any counterexample to H removes at least $\frac{1}{d+1} |S_v|$ points from S_v for some $v \in \{0, 1\}^d$ (we write $|S|$ for the number of elements of a set S). Obviously, this guarantees that altogether at most $2^d \cdot \log_{(d+1)/d} n^d = O(\log n)$ counterexamples are needed.

In order to make the idea of the learning algorithm more perspicuous, we discuss first the special case $d = 2$ (see figure 1). The general case is quite similar. For $d = 2$, the

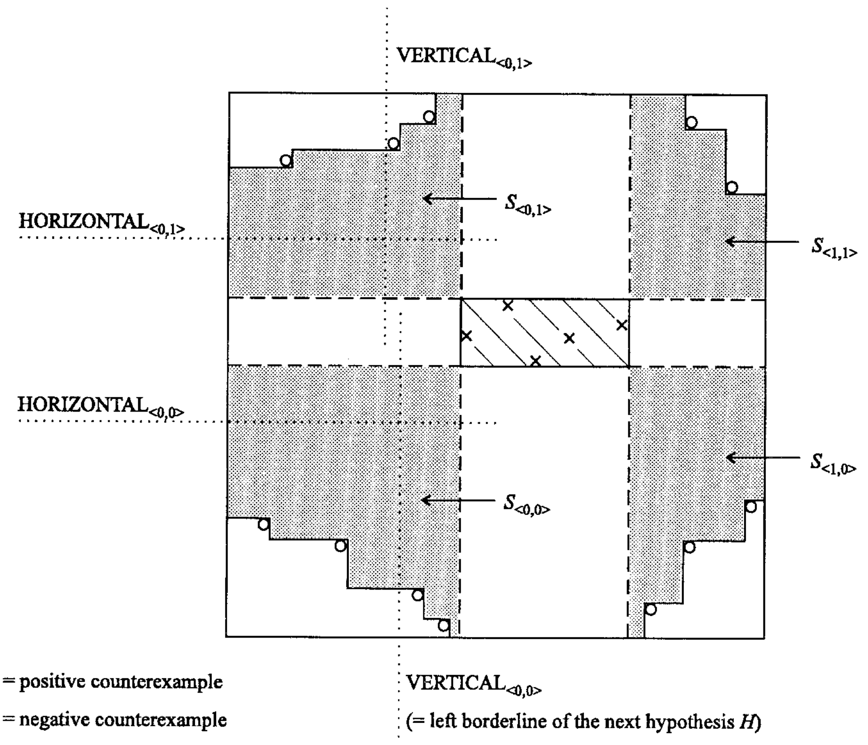
VERTICAL<0,1>
HORIZONTAL<0,1>
S<0,1>
S<1,1>
HORIZONTAL<0,0>
S<1,0>
S<0,0>
× = positive counterexample
○ = negative counterexample
VERTICAL<0,0>
(= left borderline of the next hypothesis H)

Figure 1. The special case $d = 2$.

set $S_{\langle 0,0\rangle}$ (respectively, $S_{\langle 0,1\rangle}$, $S_{\langle 1,0\rangle}$, $S_{\langle 1,1\rangle}$) consists of all points that coincide with the southwest (respectively, northwest, southeast, northeast) corner of some rectangle $C \in \mathrm{BOX}_n^2$ that is consistent with all preceding counterexamples. Let IN be the smallest $C \in \mathrm{BOX}_n^2$ that contains all preceding positive counterexamples.

The learning algorithm A for the case $d = 2$ starts with the hypothesis $H_1^A = \emptyset$. After step i, it constructs a hypothesis $H := H_{i+1}^A$ such that any counterexample to H reduces the size of at least one of the sets $S_{\langle 0,0\rangle}$, $S_{\langle 0,1\rangle}$, $S_{\langle 1,0\rangle}$, $S_{\langle 1,1\rangle}$ by one third. We fix a vertical line $\mathrm{VERTICAL}_{\langle 0,0\rangle}$ that is rightmost with the property that at least one third of the points of $S_{\langle 0,0\rangle}$ lie on or to the right of $\mathrm{VERTICAL}_{\langle 0,0\rangle}$. Then we fix a horizontal line $\mathrm{HORIZONTAL}_{\langle 0,0\rangle}$ that is high as possible with the property that at least one third of the points of $S_{\langle 0,0\rangle}$lie on or above $\mathrm{HORIZONTAL}_{\langle 0,0\rangle}$. For $S_{\langle 0,1\rangle}$, one fixes a vertical line $\mathrm{VERTICAL}_{\langle 0,1\rangle}$ that is rightmost with the property that at least one third of the points of $S_{\langle 0,1\rangle}$ lie on or to the right of $\mathrm{VERTICAL}_{\langle 0,1\rangle}$. $\mathrm{HORIZONTAL}_{\langle 0,1\rangle}$ is chosen as low as possible such that at least one third of the points of $S_{\langle 0,1\rangle}$ lie on or below it. $\mathrm{VERTICAL}_{\langle 1,1\rangle}$ is chosen leftmost such that at least one third of the points of $S_{\langle 1,1\rangle}$ lie on or to the left of it. The remaining lines are chosen in an analogous fashion.

One chooses as left borderline of the next hypothesis H the rightmost one of $\mathrm{VERTICAL}_{\langle 0,0\rangle}$, $\mathrm{VERTICAL}_{\langle 0,1\rangle}$, and as right borderline of H the leftmost one of $\mathrm{VERTICAL}_{\langle 1,0\rangle}$, $\mathrm{VERTICAL}_{\langle 1,1\rangle}$. Analogously, the upper borderline of H is determined by the lower one of the lines HORIZONTAL$_{\langle 0,1\rangle}$, HORIZONTAL$_{\langle 1,1\rangle}$, and the lower borderline of H by the higher one of HORIZONTAL$_{\langle 0,0\rangle}$, HORIZONTAL$_{\langle 1,0\rangle}$.

It is obvious that a positive counterexample that lies strictly to the left of VERTICAL$_{\langle 0,0\rangle}$ (VERTICAL$_{\langle 0,1\rangle}$) will eliminate at least one third of $S_{\langle 0,0\rangle}$ ($S_{\langle 0,1\rangle}$). Analogously, any positive counterexample that lies strictly above the upper borderline of H will eliminate at least one third of $S_{\langle 0,1\rangle}$ or $S_{\langle 1,1\rangle}$; etc. From any negative counterexample to H, one can derive for one of the corners of H (say: the $\langle 0, 0\rangle$ corner) that it does not belong to C_T. This implies that besides this $\langle 0, 0\rangle$-corner $\langle h, v\rangle$ also all points $\langle i, j\rangle \in S_{\langle 0,0\rangle}$ with $i \leq h$ and $j \leq v$ can be eliminated. Hence, all points of $S_{\langle 0,0\rangle}$ can be eliminated that lie neither strictly to the right nor strictly above this $\langle 0, 0\rangle$-corner of H. By the construction of H (and of VERTICAL$_{\langle 0,0\rangle}$, HORIZONTAL$_{\langle 0,0\rangle}$), at most (in fact: less than) one third of the points of $S_{\langle 0,0\rangle}$ lie *strictly* to the right of this $\langle 0, 0\rangle$-corner (respectively, *strictly* above this $\langle 0, 0\rangle$-corner). Hence this negative counterexample eliminates at least one third of the points of $S_{\langle 0,0\rangle}$. This finishes the sketch of the learning algorithm for the case $d = 2$.

In the general case $d \geq 2$, one also starts with the hypothesis $H_1^A = \emptyset$. After i steps (at which we have received i counterexamples), we construct the next hypothesis $H := H_{i+1}^A$ as follows. First, we define for every $v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d$ a point $h^v = \langle h_1^v, \ldots, h_d^v\rangle \in X_n^d$. If $v_k = 0$, then we choose $h_k^v \in \{0, \ldots, n - 1\}$ maximal with the property that

$$|\{\langle x_1, \ldots, x_d\rangle \in S_v \mid x_k \leq h_k^v\}| \geq \frac{1}{d + 1} |S_v|.$$

If $v_k = 1$, then we choose $h_k^v \in \{0, \ldots, n - 1\}$ minimal with the property that

$$|\{\langle x_1, \ldots, x_d\rangle \in S_v \mid x_k \leq h_k^v\}| \geq \frac{1}{d + 1} |S_v|.$$

Next, we define for every $k \in \{1, \ldots, d\}$

$$a_k^H := \max\{h_k^v \mid v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d \text{ and } v_k = 0\}$$

and

$$b_k^H := \min\{h_k^v \mid v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d \text{ and } v_k = 1\}.$$

Finally, the next hypothesis H is defined as

$$H := \mathop{\times}_{k=1}^{d} [a_k^H, b_k^H].$$

Let IN $= \times_{k=1}^d [a_k^{\text{IN}}, b_k^{\text{IN}}]$ be the smallest set $C \in \text{BOX}_n^d$ that is consistent with all of the preceding i counterexamples (such smallest box exists, because BOX_n^d is closed under intersection).

It is easy to verify that IN $\subseteq H$ (i.e., that H contains all preceding positive counterexamples): Consider an arbitrary $v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d$ and an arbitrary point $x = \langle x_1, \ldots, x_d\rangle \in S_v$. Then $x = C_v$ for some box $C = \times_{k=1}^d [a_k^C, b_k^C] \in \text{BOX}_n^d$ that is consistent with all i preceding counterexamples. By definition of IN, we have IN $\subseteq C$, and therefore

$$v_k = 0 \Rightarrow x_k = a_k^C \leq a_k^{\text{IN}},$$

and

$$v_k = 1 \Rightarrow x_k = b_k^C \geq b_k^{\text{IN}}.$$

This implies (by the definition of h_k^v) that for every $k \in \{1, \ldots, d\}$:

$$(v_k = 0 \Rightarrow h_k^v \leq a_k^{\text{IN}})$$

and

$$(v_k = 1 \Rightarrow h_k^v \geq b_k^{\text{IN}}).$$

Since the preceding inequalities hold for every $v \in \{0, 1\}^d$, they imply that for every $k \in \{1, \ldots, d\}$: $a_k^H \leq a_k^{\text{IN}}$ and $b_k^H \geq b_k^{\text{IN}}$.

Assume that $c = \langle c_1, \ldots, c_d\rangle \in H\Delta C_T$ is an arbitrary counterexample to the hypothesis H (we write C_T for the target concept). Assume first that c is a *positive* counterexample, i.e., $c \in C_T - H$. Then there exists some $k \in \{1, \ldots, d\}$ with $c_k \notin [a_k^H, b_k^H]$. Assume that $c_k < a_k^H$ (the case $c_k > b_k^H$ is analogous). By definition of a_k^H, there exists some $v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d$ with $v_k = 0$ and $a_k^H = h_k^v$. Since $c \in C_T = \times_{k=1}^d [a_k^T, b_k^T]$, we have for this v that $a_k^T \leq c_k < a_k^H = h_k^v$. Therefore, all points $\langle x_1, \ldots, x_d\rangle \in X_n^d$ with $x_k \geq h_k^v$ are eliminated from S_v (by the definition of h_d^v, this eliminates at least $(1/d + 1)|S_v|$ points from S_v).

Assume now that c is a *negative* counterexample, i.e., $c \in H - C_T$. Define $v = \langle v_1, \ldots, v_d\rangle \in \{0, 1\}^d$ as follows: if $c_K < b_k^{\text{IN}}$, set $v_k = 0$; otherwise, set $v_k = 1$. This definition of v implies immediately that the v-corner H_v of H is not in C_T. If $H_v \in C_T$, then we would get from $c \in H$ the contradiction that $c \in C_T$ (for k with $v_k = 0$, we would get that $a_k^T \leq a_k^H \leq c_k < b_k^{\text{IN}} \leq b_k^T$; for k with $v_k = 1$, we would get that $a_k^T \leq b_k^{\text{IN}} \leq c_k \leq b_k^H \leq b_k^T$).

We want to show that the negative counterexample c eliminates from S_v (for v as defined above) the set

$$\tilde{S}_v := \{\langle x_1, \ldots, x_d\rangle \in S_v \mid \forall k \in \{1, \ldots, d\}((v_k = 0 \Rightarrow x_k \leq h_k^v) \wedge (v_k = 1 \Rightarrow x_k \geq h_k^v))\}.$$

This will be sufficient, since the definition of h_k^v implies that

$$|\{\langle x_1, \ldots, x_d\rangle \in S_v \mid x_k > h_k^v\}| \leq \frac{1}{d+1} |S_v|$$

if $v_k = 0$ (by the maximality of h_k^v). Furthermore,

$$|\{\langle x_1, \ldots, x_d\rangle \in S_v \mid x_k < h_k^v\}| \leq \frac{1}{d+1} |S_v|$$

if $v_k = 1$. Thus $\tilde{S}_v$ results from S_v be subtracting from S_v d sets of size $\le 1/(d{+}1)|S_v|$; therefore, $|\tilde{S}_v| \ge 1/(d{+}1)|S_v|$. In order to show that the counterexample c eliminates from S_v all points in $\tilde{S}_v$, assume for a contradiction that the v-corner $(C_T)_v$ of the target concept C_T lies in $\tilde{S}_v$. Then we have for all k with $v_k = 0$ that

$$a_k^T \le h_k^v \le a_k^H \le a_k^{\mathrm{IN}} \le b_k^T,$$

and for all k with $v_k = 1$ that

$$b_k^T \ge h_k^v \ge b_k^H \ge b_k^{\mathrm{IN}} \ge a_k^T.$$

This yields a contradiction to the fact that $H_v \notin C_T$, which has been verified before. This completes the proof of the upper bound for theorem 1.

In order to prove that $\mathrm{LC}(\mathrm{BOX}_n^d) = \Omega(\log n)$, one uses the simple result that $\mathrm{LC}(\mathcal{C}) = \Omega(\log(\mathrm{chain}(\mathcal{C})))$ for any concept class $\mathcal{C}$ (Maass & Turán, 1989, 1992), where $\mathrm{chain}(\mathcal{C})$ is the maximal $\ell \in \mathbf{N}$ such that there exists a chain $C_1 \subsetneq C_2 \subsetneq C_2 \subsetneq \dots \subsetneq C_\ell$ of concepts in $\mathcal{C}$. It is obvious that $\mathrm{chain}(\mathrm{BOX}_n^d) \ge n$. □

3. The difficulty of learning rectangles in general position

It is shown in this section that the learning of rectangles in general position equires exponentially more learning steps than the learning of axis-parallel rectangles (for dimension $d = 2$).

We write X_n for the two-dimensional grid $\{0, \dots, n-1\}^2$. The class of rectangles in general position is defined by

$$\text{GP-BOX}_n := \{C \subseteq X_n \mid \text{there is a rectangle } R \subseteq \mathbf{R}^2 \text{ with } R \cap X_n = C \\ (R \text{ need not be axis-parallel})\}.$$

Note that $\emptyset \in \text{GP-BOX}_n$ according to this definition. It is obvious that $\mathrm{LC}(\text{GP-BOX}_n) \le |X_n| = n^2$.

Theorem 2. $\mathrm{LC}(\text{GP-BOX}_n) = \Omega(n)$.

Proof. We design an adversary strategy. The idea of the proof is to fix subsets P and N of the domain X_n so that one can apply for $\{C \cap N \mid C \in \text{GP-BOX}_n \text{ and } P \subseteq C\}$ a adversary strategy similar to the following well-known one for the concept class $\mathrm{SINGLETONS}_n := \{\{i\} \mid 1 \le i \le n\}$ (see Maass & Turán, 1989, 1992). This adversary strategy forces the learner to use $n-1$ hypotheses for learning an arbitrary target concept from $\mathrm{SINGLETONS}_n$ by responding to each hypothesis $\{i\} \in \mathrm{SINGLETONS}_n$ with the negative counterexample i. Obviously, this adversary strategy for $\mathrm{SINGLETONS}_n$ relies on the fact that $\emptyset \notin \mathrm{SINGLETONS}_n$, which prevents the learner from issuing the hypothesis $\emptyset$.

The situation for learning GP-BOX_n is different, since $\emptyset \in \text{GP-BOX}_n$. However, we choose subsets P and N of the domain X_n such that $\{C \cap N \mid C \in \text{GP-BOX}_n$ and

$P \subseteq C\}$ does *not* contain $\emptyset$. Furthermore, we make sure that one can give for any hypothesis $C \in$ GP-BOX$_n$ with $P \subseteq C$ a negative counterexample that does not eliminate too many other concepts of this type (in analogy to the situation in the adversary strategy for SINGLETONS$_n$).

We choose $P :=$ Ball $\cap X_n$ and $N :=$ Ring $\cap X_n$ for certain sets Ball, Ring that are defined below. Let dist(x, y) be the Euclidean distance between points $x, y \in \mathbf{R}^2$. We approximate the center of the domain X_n by the point $m := (\lceil n/2 \rceil, \lceil n/2 \rceil)$. The following ball and ring with center m will be considered (see also figure 2):

$$\text{Ball} := \left\{ x \in \mathbf{R}^2 \mid \text{dist}(x, m) \le \frac{n}{4} + 2 \right\},$$

$$\text{Ring} := \left\{ x \in \mathbf{R}^2 \mid \sqrt{2} \cdot \frac{n}{4} - 4 \le \text{dist}(x, m) \le \sqrt{2} \cdot \left[\frac{n}{4} + 2 \right] \right\}.$$

We assume that n is sufficiently large so that Ball $\cap$ Ring $= \emptyset$ and Ring $\subseteq \{x \in \mathbf{R} \mid 0 \le x \le n - 1\}^2$. A vertical and a horizontal line through m divided $\mathbf{R}^2$ into four quadrants. We will focus on the northwest quadrant $Q \subseteq \mathbf{R}^2$ (assume that Q contains the points on the horizontal line left of m, but no points from the vertical line).

Our adversary strategy proceeds as follows:

- If (Ball $\cap X_n) \not\subseteq H$ for the current hypothesis H, then one gives an arbitrary point from (Ball $\cap X_n) - H$ as a positive counterexample.
- If (Ball $\cap X_n) \subseteq H$, then one gives a point from $H \cap$ Ring $\cap Q$ as negative counterexample, provided that there exists some $C_T \in$ GP-BOX$_n$ that is consistent with this and all preceding counterexamples.

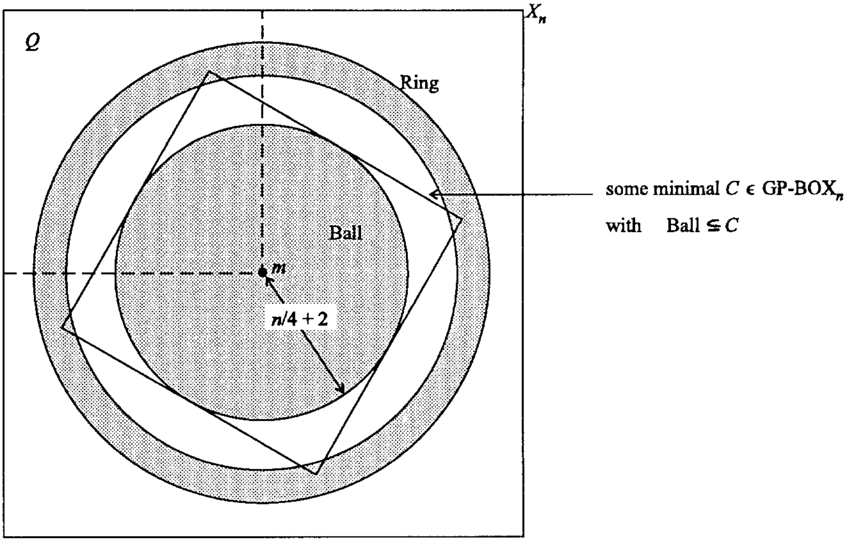

Figure 2. Ball and ring with center m.

Note that all the technical complications of this proof are caused by the fact that the considered geometrical objects (Ball, Ring, etc.) have to be intersected with the discrete domain $X_n = \{0, \ldots, n - 1\}^2$.

We will show in claim 2 that this adversary strategy is well defined (i.e., (Ball $\cap X_n) \subseteq H$ implies $H \cap$ Ring $\cap Q \neq \emptyset$). Claim 3 will imply that no learning algorithm can identify arbitrary target concepts from GP-BOX$_n$ in $o(n)$ steps if counterexamples are chosen according to this adversary strategy. Claim 1 will be needed for the proof of claim 2.

Claim 1. Consider any $z \in \{y \in \mathbf{R} \mid 1/\sqrt{2} \leq y \leq n - 1 - 1/\sqrt{2}\}^2$. Then the ball $B_z := \{x \in \mathbf{R}^2 \mid \text{dist}(x, z) \leq 1/\sqrt{2}\}$ contains some point from X_n.

Proof of claim 1. B_z contains a closed axis-parallel square with sides of length 1 that is contained in $\{x \in \mathbf{R} \mid 0 \leq x \leq n - 1\}^2$. Any such square contains a point from X_n. □

Claim 2. Assume that $H \in$ GP-BOX$_n$ and Ball $\cap X_n \subseteq H$. Then $H \cap$ Ring $\cap Q \neq \emptyset$.

Proof of claim 2. Let $R \subseteq \mathbf{R}^2$ be a rectangle with $R \cap X_n = H$. Then it need not be the case that Ball $\subseteq R$, although Ball $\cap X_n \subseteq H$. However, claim 1 implies that the slightly smaller set Ball$' := \{x \in \mathbf{R}^2 \mid \text{dist}(x, m) \leq n/4\}$ is contained in R. Otherwise, Ball $- R$ contains a ball B of radius $1/\sqrt{2}$. Claim 1 implies that $B \cap X_n \neq \emptyset$, which yields a contradiction to our assumption that Ball $\cap X_n \subseteq H = R \cap X_n$.

Since Ball$' \subseteq R$, there exists a square $R' \subseteq \mathbf{R}^2$ with sides of length $n/2$ and center m such that Ball$' \subseteq R' \subseteq R$. Obviously, R' has a corner $CO \in Q$ with dist$(CO, m) = \sqrt{2} \cdot n/4$. Let $R'' \subseteq R'$ be a square with sides of length $\sqrt{8}$ which has the same corner CO. By the definition of Ring, we have $R'' \subseteq$ Ring. Hence it is sufficient to show that $R'' \cap Q \cap X_n \neq \emptyset$. We partition R'' by the line from m to CO into two triangles. The size of R'' has been chosen large enough so that each of these two triangles contains a ball with radius $1/\sqrt{2}$. By claim 1, each of these two balls contains a point from X_n. Since $CO \in Q$, at least one of these two points from X_n lies in Q. □

Claim 3. Partition Ring $\cap Q$ into sectors by drawing from the center m in such a way that the Euclidean distance between the intersections of any two adjacent rays with the outer boundary of Ring is 60. Then any square S with center m and sides of length $n/2 + 4$ intersects Ring $\cap Q$ in at most two of these sectors (which are necessarily adjacent), provided that n is sufficiently large.

Proof of claim 3. Let CO be the corner of S that belongs to the quadrant Q. It is obvious that CO lies on the outer boundary of Ring. Pick one of the two sides of S that are incident with CO, and let A be its intersection point with the inner boundary of Ring. Let B be the point where the line $\overline{m\,A}$ intersects the outer boundary of Ring. In order to prove the claim, it is sufficient to show that dist$(CO, B) < 30$.

Let D be the point on $\overline{m\,CO}$ such that the angle $CO\ D\ A$ is orthogonal. By definition of A, the angle $A\ CO\ D$ is equal to $\pi/4$. For sufficiently large n, one has

$$\text{dist}(CO, D) \leq 2 \cdot (4 + 2 \cdot \sqrt{2}) \leq 14$$

(note that $4 + 2 \cdot \sqrt{2}$ is the difference in radius between the inner and the outer boundary of Ring). Hence,

$$\begin{aligned} \text{dist}(CO, A) &= \sqrt{2}\ \text{dist}(CO, D) \leq 14\sqrt{2}, \text{ thus} \\ \text{dist}(CO, B) &\leq \text{dist}(CO, A) + \text{dist}(A, B) \\ &\leq 14 \cdot \sqrt{2} + (4 + 2 \cdot \sqrt{2}) \leq 20 \cdot \sqrt{2} < 30. \end{aligned}$$ □

In order to complete the proof of theorem 2, we observe that the partition of claim 3 partitions Q into $\geq c \cdot n$ sectors (for some constant $c > 0$). Hence there are $\geq \lfloor cn/2 \rfloor$ disjoint pairs of adjacent sectors in Q. For each of these pairs, there exists some $C \in$ GP-BOX$_n$ with Ball $\cap X_n \subseteq C$ such that $C \cap$ Ring $\cap Q$ is contained in this pair of sectors. This observation implies that as long as not more than $\lfloor cn/2 \rfloor - 2$ negative counterexamples have been given by our adversary strategy, there are at least two different possible target concepts $C \in$ GP-BOX$_n$ that are consistent with all preceding counterexamples. Thus we have shown that LC(GP-BOX$_n$) $> \lfloor cn/2 \rfloor - 2 = \Omega(n)$. □

Remark. It is shown in Bultman and Maass (1991) that LC-MEMB(GP-BOX$_n$) $= \Theta(\log n)$. Thus rectangles in general position can be learned fast in the stronger on-line learning model where the learner can ask (besides equivalence queries with hypotheses from GP-BOX$_n$) membership queries "$x \in C_T$?" for arbitrary elements x of the domain.

4. Halfspaces, intersections of halfspaces, and semi-algebraic sets

In this section, we analyze the learning complexity of the concept classes

$$\begin{aligned} \text{HALFSPACE}_n^d := \{C \subseteq \{0, \ldots, n-1\}^d \mid & \text{ there is a halfspace} \\ & F \subseteq \mathbf{R}^d \text{ with } F \cap \{0, \ldots, n-1\}^d = C\} \\ = \Bigg\{ C \subseteq \{0, \ldots, n-1\}^d \mid & \exists w_1, \ldots, w_d, t \in \mathbf{R} \\ & \forall x_1, \ldots, x_d \in \{0, \ldots, n-1\} \\ & \left[\langle x_1, \ldots, x_d \rangle \in C \Leftrightarrow \sum_{i=1}^{d} w_i x_i \geq t \right] \Bigg\}, \end{aligned}$$

$$\text{2-HALFSPACE}_n^d := \{C \cap C' \mid C, C' \in \text{HALFSPACE}_n^d\},$$

k-SEMI-ALGEBRAIC-SETS$_n^d$:= $\{C \subseteq \{0, \ldots, n-1\}^d \mid$ there exist coefficients $w_{i_1,\ldots,i_d} \in \mathbf{R}$ for arbitrary tuples $\langle i_1, \ldots, i_d\rangle \in \mathbf{N}^d$ with $\sum_{j=1}^{d} i_j \leq k$, and there exists some $t \in \mathbf{R}$ such that for all $x_1, \ldots, x_d \in \{0, \ldots, n-1\}$

$$(\langle x_1, \ldots, x_d\rangle \in C \Rightarrow \sum_{i_1+\ldots+i_d \leq k} w_{i_1,\ldots,i_d} x_1^{i_1} \cdot \ldots \cdot x_d^{i_d} \geq t)\},$$

for an arbitrary fixed dimension $d \in \mathbf{N} - \{0\}$ and for arbitrary $k \in \mathbf{N} - \{0\}$.

It has been shown that without loss of generality, the coefficients w_i, $w_{i_1,\ldots,i_d}$ and the thresholds t in these definitions can be chosen to be integers (Muroga, 1971; see also Maass & Turán, in press).

Theorem 3. LC(HALFSPACE$_n^d$) = $\Theta(\log n)$ for every fixed dimension $d \in \mathbf{N} - \{0\}$.

Proof. It is shown in Maass and Turán (in press) that LC(HALFSPACE$_n^d$) = $O(d^2(\log d + \log n))$ and that LC(HALFSPACE$_n^d$) = $\Omega(d^2 \log n)$. □

Remarks. 1. It is an open question whether there is an *elementary* geometric construction (as for theorem 1) that shows that LC(HALFSPACE$_n^d$) = $O(\log n)$. The learning algorithm from Maass and Turán (in press) (which is computationally feasible) employs nontrivial tools from combinatorial optimization for the dual space of HALFSPACE$_n^d$.

2. The learning algorithm for HALFSPACE$_n^d$ from Maass and Turán (in press) is not necessarily consistent. However, it is very easy to make it consistent without affecting its computational feasibility or its error bound: if the algorithm maintains a list of all preceding counterexamples, then for any hypothesis H (*before* it issues H), the algorithm can check whether or not H is consistent with all preceding counterexamples. If it is not consistent with a preceding counterexample g, it does not issue hypothesis H. Instead, it proceeds (internally) as if g would be the counterexample to this hypothesis H. In this way, it moves to its next hypothesis.

3. One can also show (Maass & Turán, in press) that LC(HALFSPACE$_X^d$) = $O(d^2 \cdot \log |X|)$ for arbitrary finite domains $X \subseteq \mathbf{R}^d$, where

$$\text{HALFSPACE}_X^d := \{C \subseteq X \mid \text{there is a halfspace } H \subseteq \mathbf{R}^d \text{ with } H \cap X = C\}.$$

The learning algorithm proving this upper bound does not appear to have a computationally feasible implementation.

Theorem 4. LC(k-SEMI-ALGEBRAIC-SETS$_n^d$) = $\Theta(\log n)$ for every fixed $k, d \in \mathbf{N} - \{0\}$.

Proof. The upper bound follows from the preceding result, since one can view each product $x_1^{i_1} \cdot \ldots \cdot x_d^{i_d}$ as a new variable that ranges over $\{0, \ldots, (n-1)^d\}$. Hence one can apply a learning algorithm for HALFSPACE$^{p(k,d)}_{(n-1)^d+1}$, where $p(k, d) := |\{\langle i_1, \ldots, i_d\rangle \mid i_1, \ldots, i_d \in \mathbf{N}$ and $\Sigma_{j=1}^{d} i_j \leq k\}|$.

In order to prove the lower bound, one exhibits a chain of concepts from k-SEMI-ALGEBRAIC-SETS$_n^d$ of length n. □

In contrast to the preceding positive results, we show in the following theorem that one needs exponentially more learning steps to learn the intersection of two halfspaces (see also Maass & Turán, 1990).

Theorem 5. LC (2-HALFSPACE$_n^2$) $= \Omega(n)$.

Proof. Analogously to the proof of theorem 2, we choose sets P, $N \subseteq X_n := \{0, \ldots, n-1\}^2$ (here: $P :=$ *Square* $\cap X_n$, $N :=$ Perimeter $\cap X_n$) and apply for

$$\{C \cap N \mid C \in \text{2-HALFSPACE}_n^2 \text{ and } P \subseteq C\}$$

a similar adversary strategy as for SINGLETONS$_n$.

We define

$$\text{Square} := \left\{\left(\left\lfloor \frac{n}{2} \right\rfloor + i, \left\lfloor \frac{n}{2} \right\rfloor + j\right) \;\middle|\; i = 0, 1, j = 0, 1\right\},$$

$$\text{Perimeter} := \Big\{ (x_1, x_2) \in \mathbf{N}^2 \mid (x_1 \in \{0, n-1\} \text{ and } 0 \leq x_2 \leq n-1) \text{ or}$$
$$(x_2 \in \{0, n-1\} \text{ and } 0 \leq x_1 \leq n-1) \Big\}.$$

The adversary strategy proceeds as follows:

> If Square $\not\subseteq H$ for the current hypothesis H, then one gives an arbitrary point from Square $- H$ as positive counterexample.
>
> If Square $\subseteq H$, then one gives an arbitrary point from Perimeter $\cap H$ as a negative counterexample (provided that there exists some $C^T \in$ 2-HALFSPACE$_n^2$ that is consistent with this and all preceding counterexamples).

Claim 1. Assume C_1, $C_2 \in$ HALFSPACE$_n^2$ and Square $\subseteq C_1 \cap C_2$. Then $C_1 \cap C_2 \cap$ Perimeter $\neq \emptyset$.

Proof of claim 1. Fix a_j, b_j, $t_j \in \mathbf{R}$ such that $C_j = S_j \cap X_n$ for $S_j = \{\langle u, v\rangle \in \mathbf{R}_2 \mid a_j u + b_j v \geq t_j\}$, $j = 1, 2$. Since the convex hull of Square contains a circle of radius 1, $S_1 \cap S_2$ either contains one of the four corner points of $\{0, \ldots, n-1\}^2$, or it contains a segment of length at least 1 on one of the sides of the square determined by these four corner points. Hence, it contains a point from Perimeter. □

Claim 2. For some constant $c > 0$, there exists for every sufficiently large n a family $D_1, \ldots, D_{\lfloor cn \rfloor}$ of concepts from 2-HALFSPACE$_n^2$ such that Square $\subseteq D_i$ for $i \in \{1, \ldots, \lfloor cn \rfloor\}$ and $D_i \cap D_j \cap$ Perimeter $= \emptyset$ for every $i, j \in \{1, \ldots, \lfloor cn \rfloor\}$, $i \neq j$.

Proof of claim 2. Consider concepts $D \in$ 2-HALFSPACE$_n^2$ with Square $\subseteq D$ which are defined as the set of points in X_n between two parallel lines touching the circle that goes through the four points of Square. Clearly, there are $\Omega(n)$ concepts of this type with pairwise empty intersection on Perimeter. □

We can now complete the proof of theorem 5. It is obvious that as long as not more than $\lfloor c \cdot n \rfloor - 2$ negative counterexamples have been given according to the adversary strategy, there are at least two of the $\lfloor cn \rfloor$ concepts D_i from claim 2 that are consistent with all preceding counterexamples. This implies that

$$\mathrm{LC}(\text{2-HALFSPACE}_n^2) \geq \lfloor cn \rfloor - 1. \qquad \Box$$

5. The complexity of learning balls

For the domain $X_n^d := \{0, \ldots, n-1\}^d$, we consider the concept class

$$\mathrm{BALL}_n^d := \{C \subseteq X_n^d \mid \text{there is a ball } B \subseteq \mathbf{R}^d \text{ with } B \cap X_n^d = C\}.$$

Theorem 6. $\mathrm{LC}(\mathrm{BALL}_n^d) = O(d^2(\log d + \log n))$ and $\mathrm{LC}(\mathrm{BALL}_n^d) = \Omega(d^2 \log n)$. Furthermore, there exists a learning algorithm A for BALL_n^d with $\mathrm{LC}(A) = O(d^2(\log d + \log n))$ that uses altogether only polynomially in d and $\log n$ many computation steps.

Proof. Similarly to the case of learning semi-algebraic sets, the proof uses a reduction to halfspace learning. The class of semi-algebraic sets defined by a quadratic inequality of the form $\Sigma_{i=1}^d w_i x_i + w_{d+1}(\Sigma_{i=1}^d x_i^2) \geq t$ can be learned by introducing a new variable x_{d+1} for $\Sigma_{i=1}^d x_i^2$ and applying a learning algorithm for HALFSPACE$_{d(n-1)^2+1}^{d+1}$. The problem with this approach is that this concept class contains balls and complements of balls. If $w_{d+1} \neq 0$, then the inequality $\Sigma_{i=1}^d w_i x_i + w_{d+1}(\Sigma_{i=1}^d x_i^2) \geq t$ can be rewritten as

$$w_{d+1} \sum_{i=1}^{d} \left(x_i + \frac{w_i}{2w_{d+1}} \right)^2 \geq t + \frac{1}{4w_{d+1}} \left(\sum_{i=1}^{d} w_i^2 \right).$$

This inequality defines a ball in $\mathbf{R}^d$ iff $w_{d+1} < 0$, and the complement of a ball if $w_{d+1} > 0$. Hence a straightforward application of the halfspace learning algorithm for HALFSPACE$_{d(n-1)^2+1}^{d+1}$ would give rise to a learning algorithm for BALL_n^d that uses both balls and complements of balls as hypotheses.

In order to ensure that all hypotheses of the halfspace learning algorithm correspond to balls, we consider learning a halfspace over the extended domain

$$Y_n^{d+1} := \{0, \ldots, d(n-1)^2\}^{d+1} \cup \{v_1, v_2\}$$

in $\mathbf{R}^{d+1}$, with $v_1 := \langle 0, \ldots, 0, k\rangle$, $v_2 := \langle 0, \ldots, 0, -k\rangle$, $k := 2^{8d(\log d + \log n + 4)}$. We will show in the following two claims that balls over X_n^d correspond exactly to those halfspaces F over Y_n^{d+1} that satisfy $v_1 \notin F$ and $v_2 \in F$.

Claim 1. If $C \in \mathrm{BALL}_n^d$, then there are $w_1, \ldots, w_{d+1}, t \in \mathbf{R}$ such that

a) $C = B \cap X_n^d$ for $B = \{x \in \mathbf{R}^d \mid \Sigma_{i=1}^d w_i x_i + w_{d+1}(\Sigma_{i=1}^d x_i^2) \geq t\}$,
b) for the halfspace $F = \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i x_i \geq t\}$ it holds that $v_1 \notin F$, $v_2 \in F$.

Proof of claim 1. Let $w_1', \ldots, w_{d+1}', t' \in \mathbf{R}$ such that $C = B' \cap X_n^d$ for $B' = \{x \in \mathbf{R}^d \mid \Sigma_{i=1}^d w_i' x_i + w_{d+1}'(\Sigma_{i=1}^d x_i^2) \geq t'\}$. Consider the corresponding concept $C' = F' \cap X_{d(n-1)^2+1}^{d+1}$, where $F' := \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i' x_i \geq t\}$. Using standard bounds for the solutions of a system of linear inequalities, it follows that there are weights $w_1, \ldots, w_{d+1}$ and a threshold t such that $C' = F \cap X_{d(n-1)^2+1}^{d+1}$, for $F = \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i x_i \geq t\}$, where $w_1, \ldots, w_{d+1}, t$ are integers having absolute values less than $2^{8d(\log d + \log n + 4)}$ and $w_{d+1} < 0$. (See, e.g., Maass and Turán (in press, lemma A2) for a derivation of this bound. The condition $w_{d+1} < 0$ can be directly incorporated into the system of linear inequalities considered.) Thus, by definition, $C = B \cap X_n^d$ for the ball $B := \{x \in \mathbf{R}^d \mid \Sigma_{i=1}^d w_i x_i + w_{d+1} \cdot (\Sigma_{i=1}^d x_i)^2 \geq t\}$. Now for v_1 we have $w_{d+1} 2^{8d(\log d + \log n + 4)} \leq -2^{8d(\log d + \log n + 4)} < t$ from the bound on the absolute value of t, and hence $v_1 \notin F$ for $F := \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i x_i \geq t\}$. Similarly, $v_2 \in F$, proving claim 1. □

Claim 2. Assume that C is a halfspace over Y_n^{d+1} with $C = F \cap Y_n^{d+1}$ for some halfspace $F = \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i x_i \geq t\}$, which satisfies $v_1 \notin F$, and $v_2 \in F$. Then $w_{d+1} < 0$.

Proof of claim 2. $v_1 \notin F$ implies that $w_{d+1} k < t$, and $v_2 \in F$ implies that $w_{d+1}(-k) \geq t$. Thus one has $w_{d+1} k < w_{d+1}(-k)$, and hence $w_{d+1} < 0$. □

Now to get a learning algorithm A for BALL_n^d, assume that A^* is an algorithm for learning a halfspace over Y_n^{d+1}.

If $C_T \in \mathrm{BALL}_n^d$ is the target ball, then we simulate A^* to learn a halfspace C over Y_n^{d+1} such that for every $x = \langle x_1, \ldots, x_d\rangle \in \{0, \ldots, n-1\}^d$ it holds that $x \in C_T$ iff $\langle x_1, \ldots, x_d, \Sigma_{i=1}^d x_i^2\rangle \in C$, and furthermore $v_1 \notin C$, $v_2 \in C$. Claim 1 implies the existence of such a concept C.

If A^* presents a hypothesis H for which $v_1 \in H$ (respectively, $v_2 \notin H$), then H is not used as a hypothesis for A. Instead, one continues the simulation of A^* with v_1 (respectively, v_2) as a negative (respectively, positive) counterexample. (If both conditions hold, then the choice is arbitrary.) Otherwise, we select a halfspace $F = \{x \in \mathbf{R}^{d+1} \mid \Sigma_{i=1}^{d+1} w_i x_i \geq t\}$ such that $H = F \cap Y_n^{d+1}$. Claim 2 implies that

$$H' := \left\{ x \in \mathbf{R}^d \,\middle|\, \sum_{i=1}^{d} w_i x_i + w_{d+1} \left[\sum_{i=1}^{d} x_i^2 \right] \geq t \right\} \cap X_n^d \in \mathrm{BALL}_n^d.$$

The ball learning algorithm A presents H' as its next hypothesis. If a counterexample $x = \langle x_1, \ldots, x_d \rangle$ is received, then $\langle x_1, \ldots, x_d, \Sigma_{i=1}^d x_i^2 \rangle$ is a counterexample to H. This implies that $\mathrm{LC}(A) \leq \mathrm{LC}(A^*)$.

Hence, in order to prove the theorem, it is sufficient to prove a corresponding upper bound for $\mathrm{LC}(A^*)$. An efficient learning algorithm A^* for halfspaces over Y_n^{d+1} can be designed in the same way as an efficient learning algorithm for halfspaces over X_n^d. According to Maass and Turán (1989, in press), the latter problem can be reduced to the design of an efficient algorithm for solving the well-known convex feasibility problem in combinatorial optimization. The only difference between learning a halfspace over Y_n^{d+1} and learning a halfspace over X_n^d results from the fact that the integer coefficients of points in the domain Y_n^{d+1} are somewhat larger. This gives rise to a somewhat weaker a priori bound on the size of integer coefficients in the linear inequalities that define halfspaces over Y_n^{d+1}. In technical terms, a computation using lemmas A1 and A2 of Maass and Turán (in press) shows that one can reduce the learning of halfspaces over Y_n^{d+1} to solving the convex feasibility problem with guarantee

$$r = 2^{12(d+1)(\log(d+1)+\log n+4)}$$

(instead of $r = 2^{4d(\log d+\log n+3)}$ for the case of halfspaces over X_n^d). However, the upper bound for the query complexity of the resulting learning algorithm for halfspaces depends only on the logarithm of r ($O(\tilde{d} \log r)$ queries are needed, where $\tilde{d}$ is the dimension of the domain). Hence, there exists for halfspaces over Y_n^{d+1} a computationally feasible learning algorithm A^* with the same upper bound $O(d^2(\log d + \log n))$ on the required number of queries as for learning halfspaces over X_n^d.

The lower bound for $\mathrm{LC}(\mathrm{BALL}_n^d)$ follows by noting that $\mathrm{BALL}_n^d \supseteq \mathrm{HALFSPACE}_n^d$ and $\mathrm{LC\text{-}ARB}(\mathrm{HALFSPACE}_n^d) = \Omega(d^2 \log n)$ (Maass & Turán, in press). □

Corollary. $\mathrm{LC}(\mathrm{BALL}_n^d) = \Theta(\log n)$ for every fixed dimension $d \in \mathrm{N} - \{0\}$.

Remark. 1. It remains an open question whether the upper bound of this result can also be achieved by an elementary geometric construction (as for theorem 1).

2. Similarly to the concept class $\mathrm{HALFSPACE}_X^d$ discussed in the previous section, one can also consider the class

$$\mathrm{BALL}_X^d := \{C \subseteq X \mid \text{there is a ball } B \subseteq \mathbf{R}^d \text{ with } B \cap X = C\}$$

for an arbitrary finite domain $X \subseteq \mathbf{R}^d$. It can be shown that $\mathrm{LC}(\mathrm{BALL}_X^d) = O(d^2 \log |X|)$ for every X. Analogously to theorem 6, the learning algorithm uses a reduction to learning a halfspace over the set $X' \subseteq \mathbf{R}^{d+1}$, where

$$X' := \left\{ \left(x_1, \ldots, x_d, \sum_{i=1}^{d} x_i^2 \right) \;\middle|\; (x_1, \ldots, x_d) \in X \right\} \cup \{v_1, v_2\},$$

with $v_1 = (0, \ldots, 0, k)$, $v_2 = (0, \ldots, 0, -k)$, for some sufficiently large k.

6. Open problems

In this section we list some open problems about on-line learning of geometrical concepts. We feel that problems 1 and 5 are the most important ones.

1. Is $\text{LC}(\text{U} - 2 - \text{BOX}_n^2) = O(\log n)$ for the concept class $\text{U} - 2 - \text{BOX}_n^2 := \{C_1 \cup C_2 \mid C_1, C_2 \in \text{BOX}_n^2\}$?
2. Is $\text{LC}(\text{SQUARES}_n) = \Theta(\log n)$?
   (One defines $\text{SQUARES}_n = \{\{i_1, \ldots, j_1\} \times \{i_2, \ldots, j_2\} \mid 1 \le i_1, j_1, i_2, j_2 \le n$ and $j_1 - i_1 = j_2 - i_2\}$. It is obvious that $\text{LC}(\text{SQUARES}_n) = \Omega(\log(\text{chain}(\text{SQUARES}_n))) = \Omega(\log n)$. The best-known upper bound is $O(\log^3 n)$, due to Beals (1990).)
3. Is $\text{LC}(\text{GP-BOX}_n) = \Theta(n)$?
   (See section 3 for the lower bound $\Omega(n)$.)
4. Is $\text{LC}(\text{2-HALFSPACE}_n^2) = \Theta(n)$?
   (See section 4 for the lower bound $\Omega(n)$.)
5. Is $\text{LC}(\text{2-HALFSPACE}_2^d)$ bounded above by a polynomial in d?
   (It has been shown by Blum and Rivest (1988) that under the assumption $P \neq NP$ it is impossible that $\text{LC}(A) = O(d^{O(1)})$ for a learning algorithm A for 2-HALFSPACE_2^d, which uses only polynomially in d many *computation* steps.)

Acknowledgments

We would like to thank two anonymous referees for their helpful comments. Wolfgang Maass has been partially supported by NSF-Grant CCR 8903398, and György Turán is partially supported by OTKA-433.

References

Angluin, D. (1988). Queries and concept learning. *Machine Learning*, *2*, 319–342.

Beals, R. (1990). Unpublished manuscript.

Blum, A., & Rivest, R.L. (1988). Training a 3-node neural network is NP-complete. *Proceedings of the 1988 Workshop on Computational Learning Theory* (pp. 9–18). San Mateo, CA: Morgan Kaufmann.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M.K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, *36*, 929–965.

Bultman, W., & Maass, W. (1991). Fast identification of geometric objects with membership queries. *Proceedings of the 4th Workshop on Computational Learning Theory 1991* (pp. 337–353). San Mateo, CA: Morgan Kaufmann.

Haussler, D., Kearns, M., Littlestone, N., & Warmuth, M.K. (1991). Equivalence of models for polynomial learnability. *Information and Computation*, *95*, 129–161.

Haussler, D., Littlestone, N., & Warmuth, M.K. (1987). Expected mistake bounds for on-line learning algorithms. Unpublished manuscript.

Haussler, D., Littlestone, N., & Warmuth, M.K. (1988). Predicting {0, 1}-functions on randomly drawn points. *Proceedings of the 1st Workshop on Computational Learning Theory 1988* (pp. 280–296). San Mateo, CA: Morgan Kaufmann.

Lippmann, R.P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4–22.

Maass, W., & Turán, Gy. (1989). On the complexity of learning from counterexamples (extended abstract). *Proceedings of the 30th IEEE FOCS 1989* (pp. 262–267).

Maass, W., & Turán, Gy. (1990). On the complexity of learning from counterexamples and membership queries (extended abstract). *Proceedings of the 31st Annual IEEE FOCS* (pp. 203–210).

Maass, W., & Turán, Gy. (1992). Lower bound methods and separation results for on-line learning models. *Machine Learning*, *9*, 107–145.

Maass, W., & Turán, Gy. (In press.) How fast can a threshold gate learn? In G. Drastal, S.J. Hanson, & R. Rivest (Eds.), *Computational learning theory and natural learning systems: Constraints and prospects.* Cambridge, MA: MIT Press.

Minsky, M., & Papert, S. (1988). *Perceptrons: An introduction to computational geometry, expanded edition.* Cambridge, MA: MIT Press.

Muroga, S. (1971). *Threshold logic and its applications.* New York: Wiley.

Nilsson, N.J. (1965). *Learning machines.* New York: McGraw-Hill.

Pitt, L., & Valiant, L.G. (1988). Computational limitations on learning from examples. *Journal of the ACM*, *35*, 965–984.

Rosenblatt, F. (1962). *Principles of neurodynamics.* New York: Spartan Books.

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel distributed processing.* Cambridge, MA: MIT Press.

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

Received May 16, 1992
Accepted December 18, 1992
Final Manuscript February 18, 1993