Available online at www.sciencedirect.com

SCIENCE DIRECT®

Theoretical Computer Science 352 (2006) 85–96

Theoretical Computer Science

www.elsevier.com/locate/tcs

The DNF exception problem

Dhruv Mubayi[1], György Turán[2], Yi Zhao[*,3]

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607, USA

Received 16 July 2004; received in revised form 7 October 2005; accepted 12 October 2005

Communicated by A.A. Razborov

Abstract

Given a disjunctive normal form (DNF) expression φ and a set A of vectors satisfying the expression, called the set of exceptions, we would like to update φ to get a new DNF which is false on A, and otherwise is equivalent to φ. Is there an algorithm with running time polynomial in the number of variables, the size of the original formula and the number of exceptions, which produces an updated formula of size bounded by a certain type of function of the same parameters?

We give an efficient updating algorithm, which shows that the previously known best upper bound for the size of the updated expression is not optimal in order of magnitude. We then present a lower bound for the size of the updated formula in terms of the parameters, which is the first known lower bound for this problem. We also consider the special case (studied previously in the complexity theory of disjunctive normal forms) where the initial formula is identically true, and give efficient updating algorithms, providing new upper bounds for the size of the updated expression.

Keywords: Disjunctive normal form; Updating; Occam algorithms

1. Introduction

The problem of efficiently updating a rule in order to incorporate negative information comes up in many applications. Areas dealing with this issue from different angles include machine learning, computational learning theory, nonmonotonic logic, belief revision, databases, term rewriting and logic programming. We consider this question in the context of propositional logic, by looking at the basic case of disjunctive normal forms (DNFs).

A DNF over the variables $x_1, \ldots, x_n$ is a disjunction of terms, which are conjunctions of literals (variables or their negations). The set of vectors satisfying a DNF φ is denoted by $T(\varphi)$; the size of φ is the number of its terms. The minimal size of any DNF for a Boolean function f is denoted by $Cov(f)$.

* Corresponding author. Tel.: +1 3124133748; fax: +1 3129961491.
E-mail address: yizhao@mathstat.gsu.edu (Y. Zhao).

[1] Research partially supported by NSF Grants DMS-9970325, DMS-0400812, and an Alfred P. Sloan Research Fellowship.

[2] Also affiliated with RGAI of the Hungarian Academy of Science, Szeged. Research partially supported by NSF Grants CCR-0100036, CCF-0431059.

[3] Research partially supported by NSA Grant H98230-05-1-0079, a Campus Research Board Grant and a VIGRE Postdoctoral Fellowship at University of Illinois at Chicago.

doi:10.1016/j.tcs.2005.10.038

We think of a DNF φ as a database, representing, for example, a set of instances for which a certain property holds. Assume that the database needs to be revised, as there is a set of instances $A \subseteq T(\varphi)$, called the set of exceptions to φ, which turn out not to have the property represented by φ. The objective, then, is to compute a representation of the modified database. In other words, we would like to compute a new DNF ψ representing the Boolean function $\varphi \wedge \neg\chi_A$, where χ_A is the characteristic function of A. We refer to this computational problem as the DNF *exception problem*. (Note that besides deleting vectors from $T(\varphi)$, one could also consider the case of adding vectors to $T(\varphi)$. This can be handled in a straightforward manner by adding distinct terms for each vector, and so it is not discussed any further.)

Finding a shortest representation of ψ is easily seen to include the problem of DNF minimization, which is known to be hard to solve exactly or approximately [21,22]. Motivated by the connection of this problem to computational learning theory (see below in more detail), we are interested in a different criterion: finding efficient algorithms producing a DNF ψ of size bounded by a (possibly slowly increasing) function of the number of variables, the size of the original DNF, and the number of exceptions.

We would like to determine the function

$$XC(n, m, r) = \max\{Cov(\varphi \wedge \neg\chi_A)\},$$

where the maximum is taken over all n-variable DNFs φ with at most m terms and all $A \subseteq T(\varphi)$ with $|A| \leqslant r$. In other words, $XC(n, m, r)$ is the maximal number of terms needed in an optimal DNF for a Boolean function obtained by deleting at most r vectors from an at most m-term DNF over the variables $x_1, \ldots, x_n$. We are also interested in finding an efficient updating algorithm that produces a DNF of size close to this bound.

The DNF exception problem was first studied by Zhuravlev and Kogan [14,15,23] in the special case $m = 1$, in the context of the complexity theory of DNF. If the initial DNF consists of a single term then it can be considered to be the whole cube, and thus determining $XC(n, 1, r)$ is equivalent to determining the maximal DNF complexity of n-variable Boolean functions with r false points. In [23] it is shown that if $r = \log n - \omega(n)$, where $\omega(n) \to \infty$ arbitrarily slowly, then $XC(n, 1, r) = (1 + o(1))n$, and it is stated without proof [4] that

$$XC(n, 1, r) \leqslant \tfrac{1}{2} n \cdot r. \tag{1}$$

[14] gives lower bounds for small r, and [15] gives an upper bound of the form $O(\frac{n \cdot r}{\log n})$ for almost all sets if r is in the range $\log n < r = o(2^{n/2})$. Recently, D'yakonov [8–10] obtained new results on this problem, with emphasis on getting efficient updating algorithms for small values of r.

The general DNF exception problem was first considered by Board and Pitt [6] in their work on the relationship between PAC learnability and Occam algorithms. (An Occam algorithm is a learning algorithm that performs data compression in a certain sense. As we do not use these notions, we omit the definitions. See [13] for general background in computational learning theory, and Li et al. [17] for more recent work on this relationship.) The conjecture, implicit in [6], is that there is no upper bound for $XC(n, m, r)$ of the form

$$p(n, m, \log r) + q(n, \log m, \log r) \cdot r, \tag{2}$$

where p and q are arbitrary polynomials (in their terminology, DNF are not *strongly closed under exception lists*). If the conjecture is false, i.e., there is an upper bound of form (2) for $XC(n, m, r)$, and there is an efficient (polynomial in n, m and r) algorithm producing an updated formula of size bounded by (2), then it would follow that the existence of an Occam algorithm for DNF is necessary for the PAC learnability of DNF. It is known that the existence of an Occam algorithm is sufficient for PAC learnability in general [5]. We note that the efficient learnability of DNF, both in the PAC and the equivalence and membership query model [12], is one of the main open problems in computational learning theory.

The notion of closure under exception lists comes up in computational learning theory in a different context as well, in the work of Angluin and Kriķis [2,3] on learning with membership and equivalence queries which may contain lies. They give the general upper bound

$$XC(n, m, r) \leqslant nmr. \tag{3}$$

[4] We have been informed by a referee that Kogan's dissertation (Vych. Tsentr Akad. Nauk. SSSR, Computational Center of Academy of Science USSR, Moscow, 1987) contains a proof of (1) and a result similar to Theorem 5b.

The proof of (3) in [2,3] gives an efficient algorithm to produce an updated formula of size bounded by (3). It is implicit in a remark of Angluin and Kriķis that (1) immediately gives an improvement of (3) by a factor $\frac{1}{2}$, but for their purposes it is sufficient to have (3), which has a short proof. It may be the case that (3) is 'almost optimal'. This would imply that, indeed, DNF are not strongly closed under exception lists. As far as we know, there are no previous lower bounds for $XC(n, m, r)$.

Now we turn to the description of the results of this paper. All our positive results give upper bounds for $XC(n, m, r)$ and, at the same time, also provide efficient algorithms to produce an updated formula of size within the stated bound. In the formulation of these results, and their proofs, we only mention the upper bounds, as the algorithms themselves are then easily derived. The negative results give lower bounds for $XC(n, m, r)$.

Our first result shows that the upper bound (3) is *not* optimal up to a constant factor. The bound improves over (3) by a factor of the order $\frac{\log n}{n}$ when r is exponentially large.

Theorem 1. *Suppose that* $r \geqslant n$. *Then*

$$XC(n, m, r) \leqslant \left\lceil \frac{m}{\lfloor \log r / \log n \rfloor} \right\rceil r(n+1).$$

Now we turn to lower bound results for $XC(n, m, r)$. First, we observe the following simple linear lower bounds:

Proposition 2. (a) $XC(n, 1, 1) \geqslant n$.
(b) $XC(n, m, r) \geqslant m$ *for* $r \leqslant m \leqslant 2^{n-2}$.
(c) $XC(n, 1, r) \geqslant \frac{r}{2}$.

We obtain much stronger lower bounds in Theorems 3 and 4, where Theorem 3 Part (a) shows that (3) is in fact sharp up to a constant factor in the special case $r = 1$. We write $\binom{n}{\leqslant a} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{a}$. In the statements (and proofs) of these theorems, we omit floor and ceiling symbols.

Theorem 3. (a) *For every* $0 < \varepsilon < \frac{1}{4}$ *there is an* n_0 *such that for any* $n \geqslant n_0$ *and* $m \leqslant (e^{(2\varepsilon^2/27)n})/4$,

$$XC(n, m, 1) \geqslant \left(\tfrac{1}{4} - \varepsilon\right) nm.$$

(b) *For any* $0 < \lambda < \delta < 1$ *satisfying* $\delta^2 + 3\lambda < \delta$, *there is an* n_0 *such that for any* $n \geqslant n_0$, $m \leqslant (e^{(\lambda^2/3\delta)n})/4$ *and* $0 \leqslant \alpha < \min\{\delta^2 + 3\lambda, \delta - \delta^2 - 3\lambda\}$,

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) \geqslant m \frac{\binom{\delta}{\alpha n}}{\binom{\delta^2 n + 3\lambda n}{\alpha n}}. \tag{4}$$

After selecting appropriate λ and δ in Theorem 3 Part (b), we obtain the following lower bound.

Theorem 4. *For any* $0 < \varepsilon < \frac{1}{3}$, *there exist* $\alpha > 0$, *and* n_0 *such that*

$$XC\,(n, m, r) \geqslant m r^{1/3-\varepsilon}$$

for every $n \geqslant n_0$ *and* $m = r = \binom{n}{\leqslant \alpha n}$.

We now present improved upper bounds for the special case $m = 1$ considered in the above-mentioned papers [8–10,14,15,23]. Theorem 5(a) is a slight improvement over (1). This improvement is of interest as it is sharp for $r = n$, and thus it gives some evidence for a general conjecture for the exact value of $XC(n, 1, r)$, when r is a binomial coefficient (see Section 4). Theorem 5(b) improves the upper bound in (1) for large values of r.

Theorem 5. (a) *For $n \geqslant 2$ and $r \geqslant 2$*

$$XC(n, 1, r) \leqslant \frac{n-1}{2} r + 1.$$

(b)

$$XC(n, 1, r) \leqslant (n - \lceil \log_2 r \rceil + 1) r.$$

Note that if $r < 2^{(n+3)/2}$ then Part (a) gives a better bound, otherwise Part (b) gives a better bound.

We use the following notations in this paper. We write $\{0, 1\}^n$ for the n-dimensional hypercube. A *subcube* (or simply *cube*) of $\{0, 1\}^n$ is represented by a sequence of length n, consisting of 0's, 1's and *'s. The number of the *'s is called the *dimension* of the cube. Clearly the vectors satisfying a conjunction form a subcube, where 0's (resp. 1's) correspond to negated (resp. unnegated) variables, and *'s correspond to variables which do not occur in the conjunction, either negated or unnegated. Thus, for example, if $n = 4$, then the conjunction $x_2 \wedge \bar{x}_4$ can also be written as $*1*0$. We (ab)use the terminology of a DNF for a Boolean function, or a *cube cover* of a subset of $\{0, 1\}^n$ interchangeably, whichever is more convenient. Thus, corresponding to the definition of $Cov(f)$ in the beginning of the paper, $Cov(B)$ denotes the minimal number of cubes covering a set $B \subseteq \{0, 1\}^n$. Note that in a cube cover of $\varphi \wedge \neg\chi_A$, the cubes are not required to be disjoint, but they cannot contain any points in A. For a vector $x \in \{0, 1\}^n$, the *weight* of x is the number of ones in x. Let W_ℓ (resp. $W_{\leqslant \ell}$) be the set of vectors of $\{0, 1\}^n$ having weight ℓ (resp. at most ℓ). For a set $S \in [n] = \{1, 2, \ldots, n\}$, *the characteristic vector x_S* of S is a vector having 1 at coordinates $i \in S$ and 0 otherwise.

The structure of the rest of the paper is as follows. Section 2 gives the proofs for the general DNF exception problem and Section 3 gives the proofs for the case $m = 1$. Section 4 contains some remarks and open problems.

2. The general DNF exception problem

We prove Theorem 1 in Section 2.1, and the lower bounds, Proposition 2 and Theorems 3, 4, in Section 2.2. The proof of Theorem 1 is based on partitioning the original set of cubes into groups, replacing each group with a set of disjoint cubes, and deleting the exceptions in each group separately. The proof of Theorem 3 uses a construction which associates a Boolean function to every family of subsets of $[n]$. It is shown that if the family has a certain combinatorial property then a lower bound holds for the number of cubes required to cover the set remaining after all vectors of bounded weight are deleted. A computation providing suitable values for the constants in Theorem 3 proves Theorem 4.

2.1. Upper bound

There are standard procedures for writing a union of cubes as a disjoint union of cubes (see, e.g. [19]). These procedures are polynomial in terms of the combined size of the input and the output. Lemma 6 recalls a bound provided by these procedures. It is also known that writing a union of cubes as a disjoint union may require an exponential blowup in the number of cubes [4,7], although there is a gap between the known lower and upper bounds. We apply these procedures to a small number of cubes, so the resulting algorithm remains polynomial in the size of the input. We use $\sqcup$ to emphasize disjoint unions.

Lemma 6. (a) *For any cubes $C, C_1, \ldots, C_t$, the difference $C \setminus \bigcup_{i=1}^t C_i$ can be written as the disjoint union of at most n^t cubes.*

(b) *For any cubes $C_1, \ldots, C_t$, the union $C_1 \cup \cdots \cup C_t$ can be written as the disjoint union of at most n^t cubes.*

Proof. Part (a). By induction on t, the case $t = 1$ holds since, e.g., $\{0, 1\}^m \setminus \bar{x}_1 \wedge \cdots \wedge \bar{x}_k$ can be covered by k disjoint cubes $x_1 \sqcup (\bar{x}_1 \wedge x_2) \sqcup \cdots \sqcup (\bar{x}_1 \wedge \bar{x}_2 \wedge \cdots \wedge \bar{x}_{k-1} \wedge x_k)$. For the induction step we have

$$C \Bigg\backslash \bigcup_{i=1}^{t} C_i = (C \setminus C_t) \Bigg\backslash \bigcup_{i=1}^{t-1} C_i = \left(\bigsqcup_{j=1}^{m} D_j \right) \Bigg\backslash \bigcup_{i=1}^{t-1} C_i = \bigsqcup_{j=1}^{m} \left(D_j \Bigg\backslash \bigcup_{i=1}^{t-1} C_i \right) = \bigsqcup_{j=1}^{m} \left(\bigsqcup_{k=1}^{n_j} D_{j,k} \right),$$

where $m \leqslant n$ and $n_j \leqslant n^{t-1}$.

Part (b) follows from Part (a) after writing

$$\bigcup_{i=1}^{t} C_i = \bigsqcup_{i=1}^{t}\left(C_i \Big\backslash \bigcup_{j=1}^{i-1} C_j\right). \qquad \square$$

Lemma 7. *If $C_1, \ldots, C_s$ are disjoint cubes and $|A| \leqslant r$, then*

$$Cov\,((C_1 \cup \cdots \cup C_s) \setminus A) \leqslant s + rn.$$

Proof. Let $r_i = |C_i \cap A|$. By (3) with $m = 1$ (or Theorem 5), we have $Cov(C_i \setminus A) \leqslant nr_i$. Hence

$$Cov((C_1 \bigsqcup \cdots \bigsqcup C_s) \setminus A) \leqslant \sum_{i=1}^{s} \max(1, r_i n) \leqslant \sum_{i=1}^{s}(1 + r_i n) = s + rn. \qquad \square$$

Proof of Theorem 1. Given subcubes $C_1 \cup \cdots \cup C_m$ of $\{0, 1\}^n$, and an exception set $A \subseteq \{0, 1\}^n$ of size r, we want to show

$$Cov(C_1 \cup \cdots \cup C_m \setminus A) \leqslant \left\lceil \frac{m}{\lfloor \log r / \log n \rfloor} \right\rceil r(n+1). \tag{5}$$

We arbitrarily divide the cubes C_i into $\lceil \frac{m}{t} \rceil$ groups of at most t cubes each, where $t = \lfloor \log r / \log n \rfloor$. We then apply Lemma 6 to obtain a disjoint cube cover (of size at most n^t) of the cubes in each group. Finally when deleting A from each group, we apply Lemma 7 and get a cube cover of size $n^t + rn$. The sum of the sizes of the cube covers in all the groups is at most $\lceil \frac{m}{t} \rceil (n^t + rn)$. The bound (5) follows after substituting for t. $\square$

2.2. Lower bounds

Proof of Proposition 2. For Part (a), note that if a cube contains both $(1, 0, \ldots, 0)$ and $(0, 1, 0, \ldots, 0)$, then it must have *'s in its first two positions, and *'s or 0's in the other positions. Thus the cube must contain $\mathbf{0}$, the all-zero vector as well. Thus $Cov(\{0, 1\}^n \setminus \mathbf{0}) \geqslant n$, as each unit vector must be covered by a distinct cube. Hence $XC(n, 1, 1) \geqslant n$.

For Part (b), note that the vectors satisfying the $(n-1)$-variable parity function $x_1 \oplus \cdots \oplus x_{n-1}$ in $\{0, 1\}^n$ form 2^{n-2} one-dimensional subcubes of pairwise distance at least 2. We take m of these subcubes and remove r vectors, each from a different subcube. The remaining set still requires m cubes to be covered, thus $XC(n, m, r) \geqslant m$.

For Part (c), suppose that $2^{i-1} \leqslant r < 2^i$. We delete all vectors of even weight from a subcube C of dimension i and $r - 2^{i-1}$ vectors from $\{0, 1\}^n \setminus C$. The remaining $2^{i-1} > r/2$ vectors in C have odd weights, and each of them must be covered by a different cube. $\square$

Now we describe the combinatorial construction used in the proof of Theorem 3. For a set $S \subseteq [n]$, we denote by $Cube(S)$ the cube obtained by changing all 1's in its characteristic vector x_S to *'s. Recall that the Turán number $T(d, u, v)$ is the minimal size of a family of v-subsets of $[d]$ such that every u-subset of $[d]$ contains at least one subset from the family, where $v \leqslant u \leqslant d$ (see, e.g. [11]). Determining $T(d, u, v)$ is a major open problem in extremal combinatorics, but we only use the trivial case $T(d, u+1, 1) = d - u$ and the simple bound (see, e.g. [11])

$$T(d, u, v) \geqslant \frac{\binom{d}{v}}{\binom{u}{v}}. \tag{6}$$

Lemma 8. *Given nonnegative integers $\ell \leqslant t < d$, let $S_1, \ldots, S_m$ be subsets of $[n]$ such that $|S_i| \geqslant d$ and $|S_i \cap S_j| \leqslant t$ for every $1 \leqslant i < j \leqslant m$. Then*

$$Cov\left(\bigcup_i Cube(S_i) \setminus W_{\leqslant \ell}\right) \geqslant mT(d, t+1, \ell+1).$$

The following claim is the first step towards the proof of Lemma 8.

Proposition 9. *Let $S_1, \ldots, S_m$ be subsets of $[n]$ such that $|S_i \cap S_j| \leqslant t$ whenever $i \neq j$. If $x \in Cube(S_i)$ and $y \in Cube(S_j)$ $(i \neq j)$ are two vectors of weight $t+1$, then there is no cube $C \subseteq \bigcup_i Cube(S_i)$ containing both x and y.*

Proof. Suppose instead, that $x, y \in C$ for some cube $C \subseteq \bigcup_i Cube(S_i)$. Then, since C is a cube, the vector $x \vee y$ (the componentwise $\vee$ of x and y) also belongs to C. As $C \subseteq \bigcup_i Cube(S_i)$, there exists $1 \leqslant k \leqslant m$ such that $x \vee y \in Cube(S_k)$. Recall that $Cube(E)$ is a cube having *'s at coordinates $\ell \in E$ and 0 otherwise. Hence $Cube(S_i)$ has *'s wherever x has a 1, while $Cube(S_k)$ has *'s wherever $x \vee y$ has a 1. Therefore, $Cube(S_i)$ and $Cube(S_k)$ have at least $t+1$ (the weight of x) common *'s, or $|S_k \cap S_i| \geqslant t+1$. By assumption, this happens only if $k = i$. By repeating the same arguments to y, we obtain that $k = j$. This gives $i = j$, a contradiction. □

Proof of Lemma 8. From Proposition 9 we know that vectors of weight $t+1$ in different sets $Cube(S_i)$ have to be covered by different cubes. Let B_i denote the set of vectors of weight $t+1$ in $Cube(S_i)$, for $1 \leqslant i \leqslant m$. Lemma 8 follows if we show that covering all the points from B_i by cubes disjoint from $W_{\leqslant \ell}$ requires at least $T(d, t+1, \ell+1)$ cubes.

In fact, consider a cube C covering some points from B_i, which is disjoint from $W_{\leqslant \ell}$. Since $Cube(S_i)$ has 0's outside S_i, we may assume *w.l.o.g.* that C also has 0's outside S_i. Indeed, as C contains points from B_i, it cannot have 1's outside S_i; if it has *'s outside S_i then these can be changed to 0's without losing any points from B_i. Furthermore, C must have at least $\ell+1$ 1's, otherwise it contains a vector of weight at most ℓ. After changing some 1's to *'s if necessary, we may assume that C has exactly $\ell+1$ 1's inside S_i. Then C contains a vector $v \in Cube(S_i)$ of weight $t+1$ if and only if the set of coordinates at which C equals 1 is a subset of the set of coordinates at which v equals 1.

Therefore, the sets of size $\ell+1$ associated with the cubes covering B_i have the property that every size $t+1$ subset of S_i contains at least one of them. The minimal number of such sets is exactly the Turán number $T(d, t+1, \ell+1)$. □

The following proposition follows by simple probabilistic arguments.

Proposition 10. *Given $0 < \lambda < \delta < 1$, for sufficiently large n, if*

$$m < \min\left\{(e^{(\lambda^2/3\delta)n})/4,\ e^{(\lambda^2/6\delta^2)n}\right\}$$

then there exist (δn)-element subsets $A_1, A_2, \ldots, A_m$ such that $|A_i \cap A_j| \leqslant (\delta^2 + 3\lambda)n$ for all $i \neq j$.

Proof. Assume that n is sufficiently large. We first obtain random subsets $A'_1, \ldots, A'_m$ of $[n]$ as follows. For each $j \leqslant n$, we let j be a member of A'_i with probability δ. Thus the expectation $Exp(|A'_i|) = \delta n$ for every A'_i and $Exp(|A'_i \cap A'_j|) = \delta^2 n$ for every pair $A'_i \neq A'_j$. Let $BIN(n, p)$ denote the sum of n independent variables, each equal to 1 with probability p and 0 otherwise. The Chernoff bound (see, e.g. [1, Theorem A.11]) gives

$$Pr(|BIN(n, p) - pn| > t) < 2e^{-t^2/3pn} \quad \text{and} \quad Pr(BIN(n, p) > pn + t) < e^{-t^2/3pn}.$$

We therefore have

$$Pr(||A'_i| - \delta n| > \lambda n) < 2e^{-(\lambda^2/3\delta)n} \quad \text{and} \quad Pr(|A'_i \cap A'_j| > \delta^2 n + \lambda n) < e^{-(\lambda^2/3\delta^2)n}.$$

When $m < \min\left\{(e^{(\lambda^2/3\delta)n})/4,\ e^{(\lambda^2/6\delta^2)n}\right\}$, we have

$$2me^{-(\lambda^2/3\delta)n} + \binom{m}{2} e^{-(\lambda^2/3\delta^2)n} < \frac{1}{2} + \frac{1}{2} = 1.$$

Then there are sets $A'_1, \ldots, A'_m$ satisfying $||A'_i| - \delta n| \leqslant \lambda n$ for all i and $|A'_i \cap A'_j| \leqslant (\delta^2 + \lambda)n$ for all $i \neq j$. After adding or removing at most λn elements from each A'_i, we obtain new sets $A_1, \ldots, A_m$ of size exactly δn and for all $i \neq j$,

$$|A_i \cap A_j| \leqslant (\delta^2 + \lambda)n + \lambda n + \lambda n = \delta^2 n + 3\lambda n. \qquad \square$$

Proof of Theorem 3. To prove Part (a), given $0 < \varepsilon < \frac{1}{4}$, we apply Proposition 10 with $\delta = \frac{1}{2}$ and $\lambda = \varepsilon/3$. When $m < (e^{(\lambda^2/3\delta)n})/4 = (e^{(2\varepsilon^2/27)n})/4$, we obtain $(n/2)$-element sets $A_1, \dots, A_m$ such that $|A_i \cap A_j| \leqslant (\frac{1}{4} + \varepsilon)n$ for all $i \neq j$. Next we apply Lemma 8 to $A_1, \dots, A_m$ where $d = n/2$, $t = (\frac{1}{4} + \varepsilon)n$ and $\ell = 0$ (here we need $\varepsilon < \frac{1}{4}$ to guarantee $t < d$). It follows directly from the definition that $T(d, t+1, 1) = d - t$. Thus we conclude that

$$XC(n, m, 1) \geqslant Cov\left(\bigcup_{i=1}^{m} Cube(A_i) \setminus \mathbf{0}\right) \geqslant mT\left(\frac{n}{2}, \frac{n}{4} + \varepsilon n + 1, 1\right) = \left(\frac{1}{4} - \varepsilon\right) nm.$$

To prove Part (b), we first apply Proposition 10 and obtain (δn)-element sets $A_1, \dots, A_m$ such that $|A_i \cap A_j| \leqslant (\delta^2 + 3\lambda)n$ for all $i \neq j$. Next we apply Lemma 8 to A_i in which $d = \delta n$, $t = (\delta^2 + 3\lambda)n$ and $\ell = \alpha n$ (here we need $\alpha \leqslant \delta^2 + 3\lambda < \delta$ and n large to guarantee that $\ell \leqslant t < d$). This gives

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) \geqslant mT(\delta n, \delta^2 n + 3\lambda n + 1, \alpha n + 1).$$

The desired lower bound now follows from (6),

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) \geqslant m \frac{\binom{\delta n}{\alpha n + 1}}{\binom{\delta^2 n + 3\lambda n + 1}{\alpha n + 1}} = m \frac{\binom{\delta n}{\alpha n}}{\binom{\delta^2 n + 3\lambda n}{\alpha n}} \frac{\delta n - \alpha n}{\delta^2 n + 3\lambda n + 1} > m \frac{\binom{\delta n}{\alpha n}}{\binom{\delta^2 n + 3\lambda n}{\alpha n}},$$

where the last inequality uses the assumption that $\alpha < \delta - \delta^2 - 3\lambda$ and n is large. □

Proof of Theorem 4. Let us first recall the entropy function

$$h(x) = -x \ln x - (1-x)\ln(1-x), \quad 0 < x < 1$$

(for the convenience of later computation, we use ln, instead of $\log_2$).

Our proof consists of two steps. In Step 1, we assume that δ, λ, α satisfy the condition of Theorem 3 and in addition, $\alpha \leqslant \min\{\frac{1}{3}, (\delta^2 + 3\lambda)/2\}$. The goal of this step is to prove

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) > m \binom{n}{\leqslant \alpha n}^{\gamma_0}, \tag{7}$$

where

$$\gamma_0 = \gamma_0(\delta, \lambda, \alpha) = \frac{\delta h(\alpha/\delta) - (\delta^2 + 3\lambda) h(\alpha/(\delta^2 + 3\lambda))}{h(\alpha)}.$$

In Step 2 we will specify the values of λ, δ, α which lead to the conclusion of Theorem 4.

Step 1: Using Stirling's formula, we write

$$\binom{n}{\alpha n} = \frac{1 + o(1)}{\sqrt{2\pi\alpha(1-\alpha)n}} e^{nh(\alpha)} \quad \text{as } n \to \infty. \tag{8}$$

Applying (8) twice and using the assumption $\alpha \leqslant (\delta^2 + 3\lambda)/2$, we have

$$\frac{\binom{\delta n}{\alpha n}}{\binom{\delta^2 n + 3\lambda n}{\alpha n}} = (1 + o(1)) \sqrt{\frac{\delta(\delta^2 + 3\lambda - \alpha)}{(\delta - \alpha)(\delta^2 + 3\lambda)}} \frac{e^{\delta n h(\alpha/\delta)}}{e^{(\delta^2 + 3\lambda) n h(\alpha/(\delta^2 + 3\lambda))}} > \frac{1}{\sqrt{2}} \cdot e^{\delta n h(\alpha/\delta) - (\delta^2 + 3\lambda) n h(\alpha/(\delta^2 + 3\lambda))}.$$

Applying this to (4) and using the definition of γ_0, we have

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) > m \frac{1}{\sqrt{2}} \cdot e^{(\delta h(\alpha/\delta) - (\delta^2 + 3\lambda) h(\alpha/(\delta^2 + 3\lambda)))n} = \frac{m}{\sqrt{2}} (e^{nh(\alpha)})^{\gamma_0}.$$

For any $0 < c < 1$, we know that $xh(\frac{c}{x})$ is an increasing function for $x > c$. Therefore, $\delta^2 + 3\lambda < \delta < 1$ implies that $0 < \gamma_0 < 1$ and in particular, $(\pi\alpha(1-\alpha)n/2)^{\gamma_0} > 8$ for large n. Applying (8) again, we finally obtain

$$XC\left(n, m, \binom{n}{\leqslant \alpha n}\right) > \frac{m}{\sqrt{2}} 2\sqrt{2}\left(\frac{2}{\sqrt{2\pi\alpha(1-\alpha)n}} e^{nh(\alpha)}\right)^{\gamma_0} > 2m\binom{n}{\alpha n}^{\gamma_0} > m\binom{n}{\leqslant \alpha n}^{\gamma_0},$$

where the last inequality holds because $\gamma_0 < 1$ and $\binom{n}{\leqslant \alpha n} \leqslant 2\binom{n}{\alpha n}$ for $\alpha \leqslant \frac{1}{3}$.

Step 2: We specify the values of λ, δ, α which lead to the conclusion of Theorem 4. In order to have $\binom{n}{\leqslant \alpha n} = r = m < (e^{(\lambda^2/3\delta)n})/4$ for all large n, it suffices to have $h(\alpha) \leqslant \frac{\lambda^2}{3\delta}$ or $\lambda \geqslant \sqrt{3\delta h(\alpha)}$. We thus let $\delta = \alpha^p$, $\lambda = \sqrt{3\alpha^p h(\alpha)}$ and $\alpha > 0$ be very small. It is easy to see that all the constraints in Theorem 3 Part (b) and Step 1 hold, i.e., $\delta^2 + 3\lambda < \delta$ and $\alpha < \min\{\frac{1}{3}, (\delta^2 + 3\lambda)/2, \delta - \delta^2 - 3\lambda\}$. We claim that it suffices to have

$$\lim_{\alpha \to 0} \gamma_0\left(\alpha^p, \sqrt{3\alpha^p h(\alpha)}, \alpha\right) = p \quad \text{for every } p < 1/3. \tag{9}$$

In fact, assume (9) holds and $0 < \varepsilon < \frac{1}{3}$. We let $p = \frac{1}{3} - \varepsilon/2$ and choose $\alpha > 0$ small enough such that $\gamma_0(\alpha^p, \sqrt{3\alpha^p h(\alpha)}, \alpha) \geqslant \frac{1}{3} - \varepsilon$ and all the constraints hold. The desired bound in Theorem 4 immediately follows from (7).

To facilitate the proof of (9), we define

$$\gamma(\delta, \alpha) = \gamma_0\left(\delta, \sqrt{3\delta h(\alpha)}, \alpha\right), \quad \gamma_1(\delta, \alpha) = \gamma_0(\delta, 0, \alpha),$$

and prove the following propositions.

Proposition 11. $\lim_{\alpha \to 0} \gamma_1(\alpha^p, \alpha) - \gamma(\alpha^p, \alpha) = 0$ *for any* $p < \frac{1}{3}$.

Proposition 12. $\lim_{\alpha \to 0} \gamma_1(\alpha^p, \alpha) = p$ *for any* $p < 1$.

Proof of Proposition 11. Since

$$\gamma_1(\alpha^p, \alpha) - \gamma(\alpha^p, \alpha) = \frac{\left(\alpha^{2p} + 3\sqrt{3\alpha^p h(\alpha)}\right) h\left(\alpha/\left(\alpha^{2p} + 3\sqrt{3\alpha^p h(\alpha)}\right)\right) - \alpha^{2p} h(\alpha/\alpha^{2p})}{h(\alpha)},$$

it suffices to show that $\Gamma := (\alpha^{2p} + 3\sqrt{3\alpha^p h(\alpha)})\, h\left(\frac{\alpha}{\alpha^{2p} + 3\sqrt{3\alpha^p h(\alpha)}}\right) = \alpha^{2p} h(\alpha^{1-2p}) + \mathrm{o}(h(\alpha))$.

Using the Taylor series, we know that as $x \to 0$

$$h(x) = -x \ln x + x + \mathrm{o}(x) = -x \ln x(1 + \mathrm{o}(1)). \tag{10}$$

When $p < \frac{1}{3}$ we have $h(\alpha) = \mathrm{o}(\alpha^{3p})$ and $3\sqrt{3\alpha^p h(\alpha)} = \mathrm{o}(\alpha^{2p})$. Consequently

$$h\left(\frac{\alpha}{\alpha^{2p} + 3\sqrt{3\alpha^p h(\alpha)}}\right) = h\left(\frac{\alpha}{\alpha^{2p} + \mathrm{o}(\alpha^{2p})}\right) = h(\alpha^{1-2p} + \mathrm{o}(\alpha^{1-2p})).$$

Thus

$$\Gamma = \alpha^{2p}(1 + \mathrm{o}(1)) \cdot h(\alpha^{1-2p} + \mathrm{o}(\alpha^{1-2p})).$$

Now we need to express $h(x + \mathrm{o}(x))$ in terms of $h(x)$ (as $x \to 0$). Since $\ln x$ is continuous at $x = 1$, we have $\ln(x + \mathrm{o}(x)) = \ln x + \mathrm{o}(1)$. This and (10) imply that

$$\begin{aligned} h(x + \mathrm{o}(x)) &= -(x + \mathrm{o}(x)) \ln(x + \mathrm{o}(x)) + (x + \mathrm{o}(x)) + \mathrm{o}(x) \\ &= -(x + \mathrm{o}(x))(\ln x + \mathrm{o}(1)) + \mathrm{o}(x \ln x) \\ &= -x \ln x + \mathrm{o}(x \ln x) \\ &= h(x) + \mathrm{o}(h(x)). \end{aligned}$$

Therefore,

$$\Gamma = \alpha^{2p}(1+\mathrm{o}(1)) \cdot (h(\alpha^{1-2p}) + \mathrm{o}(h(\alpha^{1-2p}))) = \alpha^{2p} h(\alpha^{1-2p}) + \mathrm{o}(h(\alpha)),$$

where the last equality holds because $\mathrm{o}(\alpha^{2p} h(\alpha^{1-2p})) = \mathrm{o}(h(\alpha))$. $\square$

Proof of Proposition 12. Using (10), we know that as $\alpha \to 0$,

$$\begin{aligned}\alpha^p h(\alpha^{1-p}) &= \alpha^p(-\alpha^{1-p} \ln \alpha^{1-p} + \mathrm{o}(\alpha^{1-p} \ln \alpha^{1-p}))\\ &= (p-1)\alpha \ln \alpha + \mathrm{o}(\alpha \ln \alpha).\end{aligned}$$

Consequently

$$\begin{aligned}\alpha^p h(\alpha^{1-p}) - \alpha^{2p} h(\alpha^{1-2p}) &= (p-1)\alpha \ln \alpha + \mathrm{o}(\alpha \ln \alpha) - ((2p-1)\alpha \ln \alpha + \mathrm{o}(\alpha \ln \alpha))\\ &= -p\alpha \ln \alpha + \mathrm{o}(\alpha \ln \alpha)\end{aligned}$$

and

$$\gamma_1(\alpha^p, \alpha) = \frac{\alpha^p h(\alpha^{1-p}) - \alpha^{2p} h(\alpha^{1-2p})}{h(\alpha)} = \frac{-p\alpha \ln \alpha + \mathrm{o}(\alpha \ln \alpha)}{h(\alpha)} \to p \quad \text{as } \alpha \to 0. \qquad \square$$

The proof of Theorem 4 is now completed. $\square$

Remark. It seems that $\frac{1}{3}$ is actually the maximum of γ_0 in our range of the parameters. Although we do not prove this, we know that this maximum is not near 1 by the following arguments. First, because $xh(c/x)$ is an increasing function of x and $\lambda \geqslant \sqrt{3\delta h(\alpha)}$, we know that $\max \gamma_0(\delta, \lambda, \alpha) = \max \gamma(\delta, \alpha)$. Secondly,

$$\max \gamma(\delta, \alpha) \leqslant \max_{\alpha \leqslant \delta^2} \gamma_1(\delta, \alpha) = \max_{\alpha} \gamma_1\left(\sqrt{\alpha}, \alpha\right) = \max_{\alpha} \frac{\sqrt{\alpha} h\left(\sqrt{\alpha}\right)}{h(\alpha)} \leqslant \max_{0<x<1} \frac{xh(x)}{h(x^2)} = \mathcal{G},$$

where $\mathcal{G} = \frac{\sqrt{5}-1}{2} = 0.618\ldots$ is the golden ratio.

3. The exception problem for a single cube

In this section we consider the special case $m = 1$ of the exception problem. It may be assumed *w.l.o.g.* that the cube is the whole cube $\{0, 1\}^n$. We write the set A of exception vectors as $\{a_1, a_2, \ldots, a_r\} \subseteq \{0, 1\}^n$. We look for a minimal cube cover of $\{0, 1\}^n \setminus A$.

We will use two particular cubes $E_i = x_i$, and $O_i = \bar{x}_i \wedge x_{i+1}$ for $1 \leqslant i \leqslant n$ (with cyclic indexing, i.e., $O_n = \bar{x}_n \wedge x_1$). If necessary, we may emphasize the dimension n by writing E_i^n, O_i^n instead of E_i, O_i.

Let M denote the $r \times n$ matrix whose ith row is a_i. First, let us make two simple observations.

Observation 1. Switching 0's and 1's within a column, or switching two columns of M does not change the covering number.

Observation 2. If there is a column which is identically 0 or 1, meaning that A is contained in a half-cube, then a covering may be obtained by taking a minimal covering in the corresponding half-cube and adding the other half-cube. For example, suppose that the nth column of M is $\mathbf{0}$ and let a_i' be the restriction of a_i on the first $n-1$ coordinates. If $\{0, 1\}^{n-1} \setminus \{a_1', \ldots, a_r'\}$ has a cover $C_1' \cup \cdots \cup C_t'$, then $\{0, 1\}^n \setminus A$ can be covered by $C_1 \cup \cdots \cup C_t \cup E_n$, where $C_i = C_i' \wedge \bar{x}_n$.

Proof of Theorem 5, Part (a). To prove $XC(n, 1, r) \leqslant \frac{n-1}{2} r + 1$ for $r \geqslant 2$, we do induction on r and n.

The assertion trivially holds for $n = 1$. When $r = 2$, using Observation 1, we may assume that $a_1 = \mathbf{0}$ and $a_2 = 1\ldots10\ldots0$ (the first d coordinates are 1's). When $d = 1$, $E_2 \cup \cdots \cup E_n$ is a cover of size $n-1$. When $d > 1$, we first observe that $O_1^d, \ldots, O_d^d$ is a cover for $\{0, 1\}^d \setminus \{\mathbf{0}, \mathbf{1}\}$. Next, we apply Observation 2 repeatedly extending this to a cover for $\{0, 1\}^n \setminus \{a_1, a_2\}$ by adding $n-d$ additional cubes. Thus $XC(n, 1, 2) \leqslant d + n - d = n$.

Now assume $r > 2$. Let $r_j = |\{1 \leqslant i \leqslant r : a_{ij} = 0\}|$ for $j = 1, \ldots, n$.

Case 1. There exists a j_0 such that either $r_{j_0} = 0$ or $r_{j_0} = r$.

Using Observation 1 we may assume $j_0 = n$. Let a_i' denote the restriction of a_i on the first $n-1$ coordinates. By induction hypothesis (on n), there exists a cube cover of size $m \leqslant \frac{n-2}{2}r + 1$ covering $\{0,1\}^{n-1} \setminus \{a_1', \ldots, a_r'\}$. As shown in Observation 2, we may extend it to a cover for $\{0,1\}^n \setminus \{a_1, \ldots, a_r\}$ by adding one more cube. The size of the new cover $m+1$ is at most $\frac{n-2}{2}r + 2 \leqslant \frac{n-1}{2}r + 1$ (using $r \geqslant 2$).

Case 2. There exists j_0 such that $2 \leqslant r_{j_0} \leqslant r-2$.

Again assume that $j_0 = n$ and define a_i' as in Case 1. Let A_0 denote the set of vectors a_i' at which $a_{in} = 0$ and $A_1 = A - A_0$. Since $r_n \geqslant 2$ and $r - r_n \geqslant 2$, using induction hypothesis (on r), we find cubes $C_1', \ldots, C_{m_1}'$ covering $\{0,1\}^n \setminus A_0$ and $D_1', \ldots, D_{m_2}'$ covering $\{0,1\}^n \setminus A_1$, where $m_1 = (n-2)r_n/2 + 1$ and $m_2 = (n-2)(r-r_n)/2 + 1$. Therefore, $\{0,1\}^n \setminus A$ can be covered by $\bigcup_{i=1}^{m_1} C_i \cup \bigcup_{i=1}^{m_2} D_i$ with $C_i = C_i' \wedge \bar{x}_n$ and $D_i = D_i' \wedge x_n$. The size of this cover, $m_1 + m_2$ is at most $\frac{n-2}{2}r_n + 1 + \frac{n-2}{2}(r - r_n) + 1 \leqslant \frac{n-1}{2}r + 1$.

Case 3. For every $j = 1, \ldots, n$ either $r_j = 1$ or $r_j = r-1$.

Applying Observation 1, we may assume that $r_j = r-1$ for all j, i.e., each column of M has exactly one 1 and $r-1$ 0's. We consider the following two subcases.

Case 3a. $a_i \neq \mathbf{0}$ for all i.

In this case we must have $r \leqslant n$. We may assume that there exist $0 = n_0 < n_1 < \cdots < n_{r-1} < n_r = n$ which divide the integer interval $[n]$ into r blocks $B_1, B_2, \ldots, B_r$ ($B_i = [n_{i-1}+1, n_i]$) such that columns in the ith block have a 1 in the ith position:

$$
\begin{aligned}
a_1 &= \overbrace{1\cdots1}^{B_1}\ 0\cdots\ \ \cdots\ \ \cdots0\\
a_2 &= 0\cdots0\ \overbrace{1\cdots1}^{B_2}\ 0\cdots\ \ \cdots0\\
a_n &= \ 0\cdots\ \ \ \cdots\ \ \cdots0\ \overbrace{1\cdots1}^{B_r}
\end{aligned}
$$

A short case analysis (which we omit) shows that $\{0,1\}^n \setminus \{a_i\}_{i=1}^r$ is covered by the following cubes: $C_{ij} = x_{n_i+1} \wedge x_{n_j+1}$ for $0 \leqslant i < j \leqslant r-1$, $O_k = \bar{x}_k \wedge x_{k+1}$ for all $k \neq n_1, \ldots, n_r$ and $D = \bar{x}_{n_1} \wedge \bar{x}_{n_2} \wedge \cdots \wedge \bar{x}_n$. Hence we get a covering of size $\binom{r}{2} + n - r + 1 \leqslant \frac{n-1}{2}r + 1$ (using $r \geqslant 2$).

Case 3b. $a_i = \mathbf{0}$ for some i.

In this case $r \leqslant n+1$. When $r = n+1$, A consists of all the vectors of weight at most 1. Cubes $x_i \wedge x_j$, $1 \leqslant i < j \leqslant n$ give a covering of size $\binom{n}{2} < \frac{n-1}{2}(n+1) + 1$. When $r \leqslant n$, we first assume that $a_r = \mathbf{0}$. Similar to what was shown in Case 3a, there are $0 = n_0 < n_1 < \cdots < n_{r-1} = n$ such that a_i ($i < r$) has values 1 exactly in the block $B_i = [n_{i-1}+1, n_i]$. It is not hard to see that $\{0,1\}^n \setminus A$ is covered by the union of the following cubes: O_k for all $k \neq n_1, \ldots, n_{r-1}$, $D_i = x_{n_{i-1}+1} \wedge \bar{x}_{n_i}$ for $i = 1, \ldots, r-1$ and

$$
C_{ij} = *\cdots*\overbrace{1\cdots1}^{B_i}*\cdots*\overbrace{1\cdots1}^{B_j}*\cdots*,
$$

for $1 \leqslant i < j \leqslant r-1$. Hence we get a covering of size $n - (r-1) + r - 1 + \binom{r-1}{2} = \binom{r}{2} + n - r + 1 \leqslant \frac{n-1}{2}r + 1$ cubes (using $r \geqslant 2$). □

Proof of Theorem 5, Part (b). We first recall Angluin and Kriķis' proof of (3). They used a complete binary tree T of depth n to represent cubes: the root (at level 0) is $\{0,1\}^n$; for $0 \leqslant i < n$, a cube C in the ith level has left child $C \wedge x_{i+1}$ and right child $C \wedge \bar{x}_{i+1}$. In other words, the ith level consists of cubes $z_1 \wedge z_2 \wedge \cdots \wedge z_i$ with $z_i \in \{x_i, \bar{x}_i\}$ and the leaves are all vectors in $\{0,1\}^n$. Consider a set A of an exception vectors. Let T_A be the union of paths from the root to all elements in A. It is easy to see that the cubes C whose parent is contained in T_A but $C \notin T_A$ make up a cover of $\{0,1\}^n \setminus A$. The size of this cover is equal to the number of vertices with precisely one child in T_A. Clearly this number is at most $|T_A| < n|A|$, Angluin and Kriķis thus concluded that $XC(n, 1, r) \leqslant n\,r$ and consequently (3). The following lemma gives a better (in fact, tight) bound on the number of vertices with one child and Theorem 5 Part (b) thus follows. □

Lemma 13. *Suppose that T is a binary tree with r leaves, all having depth n. Then the number of vertices with one child is at most* $(n - \lceil \log r \rceil + 1)r$.

Proof of Lemma 13. Suppose that T is a tree that satisfies the hypothesis and has as many nodes with one child as possible. Let L_i be the set of vertices in its ith level and $l_i = |L_i|$. We observe that for $1 \leqslant i \leqslant n-1$, if $l_i < l_{i+1}$, then $l_{i-1} = l_i/2$. Otherwise, there must exist $v \in L_i$ and $u \in L_{i-1}$ such that v has two children w, w' and u has one child. Remove the edge (v, w'), add a new vertex v' in the ith level and add two edges (w', v') and (v', u). The new tree T' still satisfies the hypothesis and has one more vertex with one child, a contradiction. This means that $\{l_0, l_1, \dots, l_n\} = \{1, 2, 4, \dots, 2^i, r, \dots, r\}$, where $i = \lceil \log r \rceil - 1$. Consequently, the number of vertices with one child is less than $(n - i)r = (n - \lceil \log r \rceil + 1)r$. $\square$

4. Further remarks and open problems

It would be of interest to show that DNF are not strongly closed under exception lists, or to strengthen Theorem 4 by improving the exponent of r in the lower bound. The construction of Theorem 3 uses DNF containing only negated variables. It would be interesting to know, what are the best possible bounds for this class of DNF (or, equivalently, for monotone DNF).

If we delete every vector of weight k from $\{0, 1\}^n$, then every vector of weight $k-1$ and $k+1$ requires a different cube to cover it, and consequently $XC\left(n, 1, \binom{n}{k}\right) \geqslant \binom{n}{k+1} + \binom{n}{k-1}$ for $0 \leqslant k < n$. Theorem 5, Part (a) implies that this inequality becomes an equality for $k = 1$. It is easy to see that the equality holds for $k = 0$ as well. We conjecture that this is the case for every k, i.e., $XC\left(n, 1, \binom{n}{k}\right) = \binom{n}{k+1} + \binom{n}{k-1}$ for $0 \leqslant k < n$.

If the exceptions in the DNF exception problem are cubes instead of points then the number of cubes needed to cover the updated set can be exponential in the number of cubes deleted. This is shown by the example $Cov\left(\{0, 1\}^n \setminus ((\overline{x_1} \wedge \overline{x_2}) \vee \cdots \vee (\overline{x_{n-1}} \wedge \overline{x_n}))\right) = 2^{\frac{n}{2}}$.

Exception problems could also be studied over other standard machine learning domains, such as first-order logic, geometry and automata. In [20] a quantitative version is given of a special case of the seminal result of Lassez and Marriott [16] over the domain of first-order terms.

Note added in proof. In the recent paper *M. Alekhnovich, M. Braverman, V. Feldman, A. Klivans, T. Pitassi*: *Learnability and automatizability*, 45. *FOCS* (2004), 621–630 it is shown that if $NP \neq RP$ then DNF are not properly polynomially PAC learnable, solving a longstanding open problem. Their proof does not use the Occam algorithm approach discussed in the Introduction.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience [Wiley], New York, 2000.

[2] D. Angluin, M. Kriķis, Learning with malicious membership queries and exceptions, seventh Ann. ACM Conf. on Computional Learning Theory (COLT), 1994, pp. 57–66.

[3] D. Angluin, M. Kriķis, R.H. Sloan, Gy. Turán, Malicious omissions and errors in answers to membership queries, Machine Learning 28 (1997) 211–255.

[4] M.O. Ball, G.L. Nemhauser, Matroids and a reliability analysis problem, Math. Oper. Res. 4 (1979) 132–143.

[5] A. Blumer, A. Ehrenfeucht, D. Haussler, M. Warmuth, Occam's razor, Inform. Process. Lett. 24 (1987) 377–380.

[6] R. Board, L. Pitt, On the necessity of Occam algorithms, Theoret. Comput. Sci. 100 (1992) 157–184.

[7] Y. Brandman, A. Orlitsky, J. Hennessy, A spectral lower bound technique for the size of decision trees and two-level AND/OR circuits, IEEE Trans. Comput. 39 (1990) 282–287.

[8] A.G. D'yakonov, Implementation of a class of Boolean functions with a small number of zeros by irredundant disjunctive normal forms, Comput. Math. Math. Phys. 41 (2001) 775–782.

[9] A.G. D'yakonov, Test approach to the implementation of Boolean functions with few zeros by disjunctive normal forms, Comput. Math. Math. Phys. 42 (2002) 924–928.

[10] A.G. D'yakonov, Construction of disjunctive normal forms in logical recognition algorithms, Comput. Math. Math. Phys. 42 (2002) 1899–1907.

[11] Z. Füredi, Turán type problems, in: A.D. Keedwell (Ed.), Surveys in Combinatorics, London Mathematical Society Lecture Note Series, Vol. 166, Cambridge University Press, Cambridge, 1991, pp. 253–300.

[12] L. Hellerstein, V. Raghavan, Exact learning of DNF formulas using DNF hypotheses, 34 ACM STOC, 2002, pp. 465–473.
[13] M.J. Kearns, U.V. Vazirani, An Introduction to Computational Learning Theory, MIT Press, Cambridge, MA, 1994.
[14] A.Y. Kogan, On lower bounds for the complexity of disjunctive normal forms for Boolean functions with a small number of zeros, USSR Comput. Math. Math. Phys. 27 (1987) 175–181.
[15] A.Y. Kogan, Disjunctive normal forms for Boolean functions with a small number of zeros, USSR Comput. Math. Math. Phys. 27 (1987) 185–190.
[16] J.-L. Lassez, K. Marriott, Explicit representation of terms defined by counterexamples, J. Automat. Reason. 3 (1987) 301–317.
[17] M. Li, J. Tromp, P. Vitányi, Sharpening Occam's razor, Inf. Process Lett. 85 (2003) 267–274.
[19] W.G. Schneeweiss, Boolean Functions: With Engineering Applications and Computer Programs, Springer, Berlin, 1989.
[20] R.H. Sloan, I. Tsapara, Gy. Turán, On a result of Lassez and Marriott, in preparation.
[21] C. Umans, The minimum equivalent DNF problem and shortest implicants, 39 FOCS, 1998, pp. 556–563.
[22] C. Umans, Hardness of approximating Σ_2^p minimization problems, 40 FOCS, 1999, pp. 465–474.
[23] Y.I. Zhuravlev, A.Y. Kogan, Realization of Boolean functions with a small number of zeros by disjunctive normal forms, and related problems, Soviet Math. Dokl. 32 (1985) 771–775.