Exact VC-Dimension of Monotone Formulas

Ariel D. Procaccia and Jeffrey S. Rosenschein

School of Engineering and Computer Science

The Hebrew University of Jerusalem

Jerusalem, Israel

Abstract

We demonstrate that the Vapnik-Chervonenkis dimension of the class of monotone formulas over n variables is exactly $\binom{n}{\lfloor n/2 \rfloor}$.

Key words: Combinatorial problems, Computational complexity, Learnability

1 Introduction

The Vapnik-Chervonenkis (VC) dimension of a class $\mathcal{F}$ of boolean functions is a combinatorial measure of the "richness" of the class.

Definition 1. Let $\mathcal{F}$ be a class of functions $f : U \rightarrow \{0, 1\}$, let $S = \{u_1, u_2, \ldots, u_m\} \subseteq U$, and let

$$\Pi_{\mathcal{F}}(S) = \{\langle f(u_1), f(u_2), \ldots, f(u_m)\rangle : f \in \mathcal{F}\}.$$

If $|\Pi_{\mathcal{C}}(S)| = 2^m$, then S is considered *shattered* by $\mathcal{C}$.

Email address: {arielpro, jeff}@cs.huji.ac.il (Ariel D. Procaccia and Jeffrey S. Rosenschein).

Preprint submitted to Elsevier Science

Less formally, we might speak of *dichotomies* on a set S: vectors in $\{0,1\}^m$, where $|S| = m$, which assign each $u_j \in S$ a *label* in $\{0,1\}$. Similarly, we say f labels $u_j \in S$ by $b \in \{0,1\}$ if $f(u_j) = b$. Given a dichotomy on S, we say that f *realizes* the dichotomy if both labelings agree on every element of S. Using this terminology, it holds that S is shattered by $\mathcal{F}$ if $\mathcal{F}$ realizes all possible dichotomies on S.

Definition 2. The *Vapnik-Chervonenkis (VC) dimension* of $\mathcal{F}$, denoted VC-dim($\mathcal{F}$), is the size of the largest set $S \subseteq X$ that is shattered by $\mathcal{F}$. If $\mathcal{F}$ shatters arbitrarily large sets, then VC-dim($\mathcal{F}$) $= \infty$.

The VC-dimension of a class is directly related to the complexity of learning the class in the *Probably Approximately Correct (PAC)* model [7]: it yields almost matching upper [2] and lower [3] bounds. For more details, the reader is urged to consult [8].

In this paper we calculate the VC-dimension of the class of monotone formulas.

Definition 3. A *monotone formula* is a boolean formula that contains only conjunctions and disjunctions as connectives, but no negations.

In particular, a *monotone monomial* is a conjunction of literals with no negations, and a *monotone DNF formula* is a disjunction of monotone monomials. Dually, a *monotone CNF formula* is a conjunction of disjunctions of literals with no negations.

Let $\mathcal{M}_n$ be the class of all monotone formulas over the variables $\{x_1, \ldots .x_n\}$, including the constant functions *true* and $false$.[1] Each formula in $\mathcal{M}_n$ is

[1] We do not identify *true* and $false$ with 1 and 0, as the latter are reserved for labelings.

regarded as a function into $\{true, false\}$ whose domain is the set of all possible assignments to $\{x_1, \ldots, x_n\}$: the formula labels an assignment by 1 iff the assignment satisfies the formula.

The VC-dimension of some subclasses of $\mathcal{M}_n$ has previously been investigated. Natschläger and Schmitt [5] have proven that the VC-dimension of the class of monotone monomials over n variables is n for all $n \in \mathbb{N}$. Bounds on the VC-dimension of k-term monotone DNF formulas have been established in [4,6].

In the following we show that the VC-dimension of the class $\mathcal{M}_n$ of monotone formulas over n variables is $\binom{n}{\lfloor n/2 \rfloor}$.[2]

2 Bounds

The lower bound is quite straightforward.

Lemma 1. $\forall n \in \mathbb{N}$, VC-dim$(\mathcal{M}_n) \geq \binom{n}{\lfloor n/2 \rfloor}$.

Proof. We must show that there is a set S of assignments to the variables $\{x_1, \ldots, x_n\}$, such that $|S| = \binom{n}{\lfloor n/2 \rfloor}$, which is shattered by $\mathcal{M}_n$. Consider the set S of all assignments A which assign *true* to exactly $\lfloor n/2 \rfloor$ of the variables; it holds that $|S| = \binom{n}{\lfloor n/2 \rfloor}$. Fix some dichotomy on S, and let $T = \{A \in S : A \text{ is labeled by } 1\} \subseteq S$ be the set of all assignments in S which the dichotomy labels by 1. It suffices to show that there is a monotone formula which is satisfied by all assignments in T, and not satisfied by any of the assignments in $S \setminus T$, as this implies that an arbitrary dichotomy can be realized by $\mathcal{M}_n$.

[2] Recall that for $n = 1$, $\binom{1}{0} = 1$.

Observe the monotone formula

$$f(x_1,\ldots,x_n) = \bigvee_{A\in T} \bigwedge_{i:\ A(x_i)=true} x_i. \tag{1}$$

We first consider some special cases. If the dichotomy labels all assignments in S by 0, then $T=\emptyset$, and thus f is an empty disjunction — the constant function $false$; this formula is not satisfied by any assignment, hence f labels all assignments in S by 0, and the dichotomy is realized. The only case where the disjunction in Equation 1 is not empty, but the conjunction is empty, is the one where $n=1$, and the single assignment in S is labeled by 1. In this case, we have that f is an empty conjunction, and thus f is the constant function $true$, so the dichotomy is again realized.

Consequently, we may assume hereinafter that the disjunction and the conjunction in Equation (1) are both nonempty. For all $A\in T$, it holds that f is satisfied by A, as the monomial $\bigwedge_{i:\ A(x_i)=true} x_i$ is satisfied by A. Moreover, f is not satisfied by any of the assignments $A'\in S\setminus T$. Indeed, all monomials in f are conjunctions of $\lfloor n/2\rfloor$ variables, and A' assigns $true$ to exactly $\lfloor n/2\rfloor$ variables; thus, in each monomial there is at least one variable which is assigned $false$ by A'. $\square$

In order to establish an upper bound, we first need to address a natural combinatorial problem. Given the numbers $\{1,2,\ldots,n\}$, we would like to find a maximal *antichain* of subsets, i.e., a family of subsets such that for any two subsets, neither one is contained in the other. Finding an antichain of size $\binom{n}{\lfloor n/2\rfloor}$ is easy: we simply choose all subsets of size $\lfloor n/2\rfloor$. But can one do better? Sperner's Theorem [1] gives a negative answer.

Theorem 1 (Sperner)**.** *Let $\mathcal{F}$ be a family of subsets of $\{1,2,\ldots,n\}$, such that* *for all* $A, B \in \mathcal{F}$*:* $A \nsubseteq B$*. Then* $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

Lemma 2. $\forall n \in \mathbb{N},\ \text{VC-dim}(\mathcal{M}_n) \leq \binom{n}{\lfloor n/2 \rfloor}$.

Proof. We need to prove that every set S of assignments such that $|S| > \binom{n}{\lfloor n/2 \rfloor}$ cannot be shattered by $\mathcal{M}_n$. Let S be such a set, and for each $A \in S$, let $V_A = \{i \in \{1, \ldots, n\} :\ A(x_i) = true\}$. By Sperner's Theorem, there must be two assignments $A_1 \neq A_2$ in S such that $V_{A_1} \subseteq V_{A_2}$. To put it differently, there must be two different assignments A_1 and A_2 in S such that for all $i = 1, \ldots n$:

$$[A_1(x_i) = true] \Rightarrow [A_2(x_i) = true]. \tag{2}$$

Let $f \in \mathcal{M}_n$ which is satisfied by A_1 (f labels A_1 by 1). It follows from equation (2) and the monotonicity of f that f is also satisfied by A_2, or in other words — f labels A_2 by 1. Therefore, any dichotomy which labels A_1 by 1 and A_2 by 0 cannot be realized by $\mathcal{M}_n$. □

3 Conclusions

By combining Lemmas 1 and 2 we obtain:

Theorem 2. $\forall n \in \mathbb{N},\ \text{VC-dim}(\mathcal{M}_n) = \binom{n}{\lfloor n/2 \rfloor}$

Notice that any monotone formula can be transformed into a monotone CNF or DNF formula by repeatedly using the distributivity of $\wedge, \vee$. Therefore, for all n, the VC-dimension of the class of monotone CNF/DNF formulas over n variables is also n.

References

[1] V. K. Balakrishnan. *Theory and Problems of Combinatorics*. Schaum's Outline Series. McGraw-Hill, 1995.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[3] A. Ehrenfeucht, D. Haussler, M. Kearns, and L. Valiant. A general lower bound on the number of examples needed for learning. *Information and Computation*, 82(3):247–261, 1989.

[4] N. Littlestone. Learning quickly when irrelevant attributes abound. *Machine Learning*, 2(4):285–318, 1988.

[5] T. Natschläger and M. Schmitt. Exact VC-dimension of boolean monomials. *Information Processing Letters*, 59:19–20, 1996.

[6] M. Schmitt. On the capabilities of higher-order neurons: A radial basis function approach. *Neural Computation*, 17(3):715–729, 2005.

[7] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[8] M. Vidyasagar. *A Theory of Learning and Generalization*. Springer-Verlag, 1997.