Exact VC-Dimension of Boolean Monomials*

Thomas Natschläger Michael Schmitt

Institute for Theoretical Computer Science
Technische Universität Graz
Klosterwiesgasse 32/2
A-8010 Graz, Austria
{tnatschl,mschmitt}@igi.tu-graz.ac.at

Abstract

We show that the Vapnik-Chervonenkis dimension of Boolean monomials over n variables is at most n for all $n \geq 2$. It follows that the VC-dimension is determined exactly and is, except for $n = 1$, equal to the VC-dimension of the proper subclass of monotone monomials.

KEYWORDS: Combinatorial Problems, Computational Complexity, Learnability

*Work supported by the ESPRIT Working Group NeuroCOLT No. 8556

1 Introduction

The *Vapnik-Chervonenkis dimension* VC-dim($\mathcal{C}$) of a collection $\mathcal{C}$ of subsets of a set X is defined as the maximum cardinality of any set $S \subseteq X$ that is shattered by $\mathcal{C}$. A set S is *shattered by* $\mathcal{C}$ if for every subset T of S there exists a $C \in \mathcal{C}$ such that $T = S \cap C$.

The VC-dimension of a class $\mathcal{F}$ of functions $f : X \to \{0,1\}$ is defined by identifying each element $f \in \mathcal{F}$ with the set $\{x \in X : f(x) = 1\}$. Thus VC-dim($\mathcal{F}$) is the maximum cardinality of any set $S \subseteq X$ for which $\mathcal{F}$ induces all functions $g : S \to \{0,1\}$.

The Vapnik-Chervonenkis dimension gives almost tight bounds for the number of examples required for learning in Valiant's PAC-model [5]. For a detailed description we refer the reader to the article of Blumer *et al.* [2] and to the books of Anthony and Biggs [1] and Natarajan [4].

In this paper we investigate the VC-dimension of *Boolean monomials*. The class MONOMIALS$_n$ is the set of all conjunctions of literals over the variables $\{x_1, \ldots, x_n\}$, including the constant functions **0** and **1**. A monomial is called *monotone* if it does not contain negations. The corresponding class is denoted by MONOTONE-MONOMIALS$_n$. We also include the constant functions **0** and **1** in the class MONOTONE-MONOMIALS$_n$.

The upper bound $(\log 3)n$ for VC-dim(MONOMIALS$_n$) has been given by Anthony and Biggs [1, p. 76] based on the familiar relationship VC-dim($\mathcal{F}$) $\leq \log|\mathcal{F}|$ for finite $\mathcal{F}$.[1] Here, log denotes the logarithm to base 2. A lower bound of n has been known for quite a while [3]. The latter proof in fact uses monotone monomials only, hence VC-dim(MONOTONE-MONOMIALS$_n$) $\geq n$.

In the following we show that n is also an upper bound even for the class MONOMIALS$_n$ for $n \geq 2$. Thus, the VC-dimension of monomials is determined exactly. Furthermore, it follows that adding negations to monotone monomials does not increase the VC-dimension, except for $n = 1$. The results are easily transferred to the dual class of Boolean clauses.

2 The upper bound

Theorem 2.1 *VC-dim(MONOMIALS$_n$) $\leq n$ for all $n \geq 2$.*

Proof. Let $S \subseteq \{0,1\}^n$ be an arbitrary set of cardinality $n+1$ and assume that it can be shattered by MONOMIALS$_n$. We fix an enumeration $u^{(1)}, \ldots, u^{(n+1)}$ of the elements of S and define $S_i := S \setminus \{u^{(i)}\}$ for $i = 1, \ldots, n+1$. The definition of shattering implies in particular that for each S_i there exists a monomial $m_i \in$ MONOMIALS$_n$ such that $S_i = S \cap m_i$. Thus, for $i, j = 1, \ldots, n+1$

[1] The authors of [1] disregard the function **0** in their definition of monomials (see p. 12). Thus, the bound $\log(3^n + 1)$ is more advisable in our case.

$$m_i \quad \text{is false on} \quad u^{(j)} \qquad \text{iff} \qquad i = j. \tag{1}$$

Therefore, each $u^{(i)}$ must contain a component $u^{(i)}_{h(i)}$ and each m_i must contain a literal $l_{k(i)}$ such that $l_{k(i)}$ is false on $u^{(i)}_{h(i)}$. Among the literals $l_{k(1)}, \ldots, l_{k(n+1)}$ at least one variable occurs twice. Without loss of generality we assume that $l_{k(1)}$ and $l_{k(2)}$ both contain the same variable. Then there are two cases: $l_{k(1)} = l_{k(2)}$ and $l_{k(1)} = \neg l_{k(2)}$.

If $l_{k(1)} = l_{k(2)}$ then $l_{k(1)}$ is false on both $u^{(1)}_{h(1)}$ and $u^{(2)}_{h(2)}$. But then m_1 is false on both $u^{(1)}$ and $u^{(2)}$ in contradiction to (1).

If $l_{k(1)} = \neg l_{k(2)}$ then consider $u^{(3)}$. (Recall that $n \geq 2$.) Either $l_{k(1)}$ or $l_{k(2)}$ is false on $u^{(3)}$. Consequently, either m_1 is false on $u^{(3)}$ or m_2 is false on $u^{(3)}$ in contradiction to (1). □

It is easy to see that the set $\{0, 1\}$ can be shattered by MONOMIALS$_1$. Therefore, we have VC-dim(MONOMIALS$_1$) = 2.

3 Conclusions

Together with the lower bound established by Ehrenfeucht *et al.* [3] we obtain precise values for the VC-dimension of monomials:

Corollary 3.1 *VC-dim(MONOMIALS$_n$)* $= \begin{cases} n & \text{if} \quad n \geq 2 \\ 2 & \text{if} \quad n = 1. \end{cases}$

The cited lower bound result also holds for monotone monomials. The set $\{0, 1\}$ cannot be shattered by MONOTONE-MONOMIALS$_1$ because this class contains only three functions $\{\mathbf{0}, \mathbf{1}, x\}$. Therefore, the VC-dimensions of both classes are equal except for $n = 1$.

Corollary 3.2 *VC-dim(MONOTONE-MONOMIALS$_n$) $= n$ for all n.*

Finally, we state that the results are transferable to MONOTONE-CLAUSES$_n$ and CLAUSES$_n$ by duality. A (monotone) clause is a disjunction of (non-negated) literals.

Corollary 3.3 *VC-dim(CLAUSES$_n$)* $= \begin{cases} n & \text{if} \quad n \geq 2 \\ 2 & \text{if} \quad n = 1 \end{cases}$
and VC-dim(MONOTONE-CLAUSES$_n$) $= n$ for all n.

References

[1] Martin Anthony and Norman Biggs. *Computational Learning Theory*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 1992.

[2] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the Association for Computing Machinery*, 36:929–965, 1989.

[3] Andrzej Ehrenfeucht, David Haussler, Michael Kearns, and Leslie Valiant. A general lower bound on the number of examples needed for learning. *Information and Computation*, 82:247–261, 1989.

[4] Balas K. Natarajan. *Machine Learning: A Theoretical Approach*. Morgan Kaufmann, San Mateo, CA, 1991.

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