Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 156 (2008) 2753–2767
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Constrained versions of Sauer's lemma

Joel Ratsaby

Department of Electrical and Electronic Engineering, Ariel University Center of Samaria, Ariel 40700, Israel

Received 8 July 2006; received in revised form 10 November 2007; accepted 18 November 2007
Available online 3 January 2008

Abstract

Let $[n] = \{1, \dots, n\}$. For a function $h : [n] \to \{0, 1\}$, $x \in [n]$ and $y \in \{0, 1\}$ define by the *width* $\omega_h(x, y)$ of h at x the largest nonnegative integer a such that $h(z) = y$ on $x - a \le z \le x + a$. We consider finite VC-dimension classes of functions h constrained to have a width $\omega_h(x_i, y_i)$ which is larger than N for all points in a sample $\zeta = \{(x_i, y_i)\}_1^\ell$ or a width no larger than N over the whole domain $[n]$. Extending Sauer's lemma, a tight upper bound with closed-form estimates is obtained on the cardinality of several such classes.

Keywords: Sauer's lemma; Integer partitions; Binary functions

1. Introduction

Let $[n] = \{1, \dots, n\}$ and denote by $2^{[n]}$ the class of all 2^n functions $h : [n] \to \{0, 1\}$. Let $\mathcal{H}$ be a class of functions and for a set $A = \{x_1, \dots, x_k\} \subseteq [n]$ denote by $h_{|A} = [h(x_1), \dots, h(x_k)]$ the *restriction* of h on A. A class $\mathcal{H}$ is said to *shatter* A if $\left|\{h_{|A} : h \in \mathcal{H}\}\right| = 2^k$. The Vapnik–Chervonenkis dimension of $\mathcal{H}$, denoted as $VC(\mathcal{H})$, is defined as the cardinality of the largest set shattered by $\mathcal{H}$. The following well-known result obtained by [19,21,24] states a tight upper bound on the cardinality of classes $\mathcal{H}$ of VC-dimension d.

Lemma 1 (*Sauer's Lemma*). *For any* $1 \le d < n$ *let*

$$\mathbb{S}(n, d) = \sum_{k=0}^{d} \binom{n}{k}.$$

Then

$$\max_{\mathcal{H} \subset 2^{[n]} : VC(\mathcal{H}) = d} |\mathcal{H}| = \mathbb{S}(n, d).$$

More generally, the lemma holds for classes of finite VC-dimension on infinite domains.

E-mail address: ratsaby@ariel.ac.il.
URL: http://www.ariel.ac.il/ee/pf/ratsaby.

0166-218X/$ - see front matter
doi:10.1016/j.dam.2007.11.017

y					$y_1=1$									$y_2=0$							
h_1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
h_2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0
[n]	1	2	.	.	x_1	.	.	.	.	.	.	.	.	x_2	.	.	.	.	.	.	n

Fig. 1. $\omega_{h_1}(\zeta) = \omega_{h_2}(\zeta) = 3$.

Aside of being an interesting combinatorial result (see Chapter 17 in [9]), Lemma 1 has been instrumental in statistical learning theory [23], combinatorial geometry [17], graph theory [4,16] and in the theory of empirical processes [18]. In such areas, the complexity of analysis of algorithms on discrete structures, for instance, learning an unknown target binary function, typically involves a simpler structure constrained by some 'smoothness' property which is induced by the underlying algorithmics. In learning, the constraint is induced by a finite sample.

Consider a binary function $h : [n] \to \{0, 1\}$, $x \in [n]$ and $y \in \{0, 1\}$ and define by $\omega_h(x, y)$ the largest a, $0 \le a \le \min\{x, n - x\}$ such that $h(z) = y$ for all $x - a \le z \le x + a$; if no such a exists then let $\omega_h(x, y) = -1$. We call this the *width* of h at x with respect to y. Denote by $\Xi = [n] \times \{0, 1\}$. By a *sample* $\zeta = \{(x_i, y_i)\}_{i=1}^{\ell} \in \Xi^{\ell}$, we mean a set of ℓ pairs with different x-components. Define by $\omega_h(\zeta) = \min_{1\le i\le \ell} \omega_h(x_i, y_i)$ the width of h with respect to ζ. For instance, Fig. 1 displays a sample $\zeta = \{(x_1, y_1), (x_2, y_2)\}$ and two functions h_1, h_2 which have a width of 3 with respect to ζ. In [3], the complexity of learning binary functions by aiming to maximize this sample width has been investigated.

The main question posed in this paper is as follows: starting from a class $\mathcal{H}$ as above with $\mathrm{VC}(\mathcal{H}) = d$ consider a subset of $\mathcal{H}$ of functions which are 'smooth', i.e., constrained to have large sample widths and therefore consecutive runs of 1's or runs of 0's of a certain minimal length. Does Sauer's lemma hold for such a subset? How does its cardinality increase with respect to n and how is it affected by the size of the allowed sample width?

The area of research on Poisson approximations (see for instance [5–7]) includes many results on the number of binary sequences of length n that have 'long' repetitive runs (with various definitions of a long run). Our question above differs in that we add the condition of having a known VC-dimension. To our knowledge, this is the first instance of a study which considers estimating the complexity of a class constrained structurally by both an extremal set property (having a finite VC-dimension) and a repetitive-run type property.

Let $N \ge 0$ be a width parameter. We study the complexity of classes of the form

$$\mathcal{H}_N(\zeta) = \{h \in \mathcal{H} : \omega_h(\zeta) > N\}, \quad \mathrm{VC}(\mathcal{H}) = d \tag{1}$$

where $\zeta = \{(x_i, y_i)\}_{i=1}^{\ell} \in \Xi^{\ell}$ is a given sample.

We obtain tight bounds in the form of Sauer's Lemma 1 on the cardinality of such classes. It turns out that the bounds have subtle nonlinear dependence on n and N. This is investigated in detail in subsequent sections.

For a function $h : [n] \to \{0, 1\}$ let the *difference* function be defined as

$$\delta_h(x) = \begin{cases} 1 & \text{if } h(x-1) = h(x) \\ 0 & \text{otherwise} \end{cases}$$

where we assume that any h satisfies $h(0) = 0$ (see Fig. 2). Define

$$\mathcal{D}_{\mathcal{H}} \equiv \{\delta_h : h \in \mathcal{H}\}, \tag{2}$$

or $\mathcal{D}$ for brevity. It is easy to see that the class $\mathcal{D}$ is in one-to-one correspondence with $\mathcal{H}$. For $N \ge 0$ and any sample ζ, if $\omega_h(x, y) \le N$ for $(x, y) \in \zeta$ then the corresponding δ_h has $\omega_{\delta_h}(x, 1) \le N$. In order to estimate the cardinality of classes $\mathcal{H}_N(\zeta)$ we will estimate the cardinality of the corresponding difference classes $\mathcal{D}_N(\zeta_+)$ which are defined based on $\zeta_+ = \{(x_i, 1) : (x_i, y_i) \in \zeta, 1 \le i \le l\}$. We denote by

$$\mathrm{VC}_{\Delta}(\mathcal{H}) \equiv \mathrm{VC}(\mathcal{D})$$

the VC-dimension of the difference class $\mathcal{D} = \{\delta_h : h \in \mathcal{H}\}$ and use it to characterize the complexity of $\mathcal{H}$ (it is easy to show that $\mathrm{VC}(\mathcal{D}) \le c\mathrm{VC}(\mathcal{H})$ for some small constant $c > 1$). We henceforth denote by $d \equiv \mathrm{VC}_{\Delta}(\mathcal{H})$.

The rest of the paper is organized as follows: in Section 2 we state the main results, Section 3 contains the lemmas used for proving these results.