IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 103–107

DOI: 10.1214/09-IMSCOLL508

A note on bounds for VC dimensions

Aad van der Vaart[1] **and Jon A. Wellner**[2,*]

Vrije Universiteit, Amsterdam and University of Washington, Seattle

Abstract: We provide bounds for the VC dimension of class of sets formed by unions, intersections, and products of VC classes of sets $\mathcal{C}_1, \ldots, \mathcal{C}_m$.

1. Introduction and main results

Let $\mathcal{C}$ be a class of subsets of a set $\mathcal{X}$. An arbitrary set of n points $\{x_1, \ldots, x_n\}$ has 2^n subsets. We say that $\mathcal{C}$ *picks out* a certain subset from $\{x_1, \ldots, x_n\}$ if this can be formed as a set of the form $C \cap \{x_1, \ldots, x_n\}$ for some $C \in \mathcal{C}$. The collection $\mathcal{C}$ is said to *shatter* $\{x_1, \ldots, x_n\}$ if each of its 2^n subsets can be picked out by $\mathcal{C}$. The *VC - dimension* $V(\mathcal{C})$ is the largest cardinality of a set shattered by $\mathcal{C}$ (or $+\infty$ if arbitrarily large finite sets are shattered); more formally, if

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\{C \cap \{x_1, \ldots, x_n\} : \ C \in \mathcal{C}\},$$

then

$$V(\mathcal{C}) = \sup\Big\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) = 2^n\Big\},$$

and $V(\mathcal{C}) = -1$ if $\mathcal{C}$ is empty. (The VC-dimension $V(\mathcal{C})$ defined here corresponds to $S(\mathcal{C})$ as defined by [5] page 134. Dudley, and following him ourselves in [11], used the notation $V(\mathcal{C})$ for the *VC-index*, which is the dimension plus 1. We have switched to using $V(\mathcal{C})$ for the VC-dimension rather than the VC-index, because formulas are simpler in terms of dimension and because the machine learning literature uses dimension rather than index.)

Now suppose that $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$ are VC-classes of subsets of a given set $\mathcal{X}$ with VC dimensions $V_1, \ldots, V_m$. It is known that the classes $\sqcup_{j=1}^m \mathcal{C}_j$, $\sqcap_{j=1}^m \mathcal{C}_j$ defined by

$$\begin{aligned} \sqcup_{j=1}^m \mathcal{C}_j &\equiv \{\cup_{j=1}^m C_j : \ C_j \in \mathcal{C}_j, \ j = 1, \ldots, m\}, \\ \sqcap_{j=1}^m \mathcal{C}_j &\equiv \{\cap_{j=1}^m C_j : \ C_j \in \mathcal{C}_j, \ j = 1, \ldots, m\}, \end{aligned}$$

are again VC: when $\mathcal{C}_1 = \cdots = \mathcal{C}_m = \mathcal{C}$ and $m = k$, this is due to [2] (see also [3], Theorem 9.2.3, page 85, and [5], Theorem 4.2.4, page 141); for general $\mathcal{C}_1$, $\mathcal{C}_2$ and $m = 2$ it was shown by [3], Theorem 9.2.6, page 87, (see also [5], Theorem 4.5.3, page 153), and [9], Lemma 15, page 18. See also [8], Lemma 2.5, page 1032. For a summary of these types of VC preservation results, see e.g. [11], page 147. Similarly,

*Supported in part by NSF Grant DMS-0804587, and by NI-AID grant 2R01 AI291968-04.
[1]Department of Mathematics, Faculty of Sciences, Vrije Universiteit De Boelelaan 1081a, 1081 HV Amsterdam, e-mail: aad@cs.vu.nl
[2]Department of Statistics, University of Washington, Seattle, WA 98195-4322, e-mail: jaw@stat.washington.edu
AMS 2000 subject classifications: Primary 60B99; secondary 62G30.
Keywords and phrases: Vapnik-Chervonenkis class, combining classes, inequality, entropy.

if $\mathcal{D}_1, \ldots, \mathcal{D}_m$ are VC-classes of subsets of sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, then the class of product sets $\boxtimes_{j=1}^m \mathcal{D}_j$ defined by

$$\boxtimes_{j=1}^m \mathcal{D}_j \equiv \{D_1 \times \ldots \times D_m : \ D_j \in \mathcal{D}_j, \ j = 1, \ldots, m\}$$

is a VC-class of subsets of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. This was proved in [1], Proposition 2.5, and in [3], Theorem 9.2.6, page 87 (see also [5], Theorem 4.2.4, page 141).

In the case of $m = 2$, consider the maximal VC dimensions $\max V(\mathcal{C}_1 \sqcup \mathcal{C}_2)$, $\max V(\mathcal{C}_1 \sqcap \mathcal{C}_2)$, and $\max V(\mathcal{D}_1 \boxtimes \mathcal{D}_2)$, where the maxima are over all classes $\mathcal{C}_1, \mathcal{C}_2$ (or $\mathcal{D}_1, \mathcal{D}_2$ in the last case) with $V(\mathcal{C}_1) = V_1$, $V(\mathcal{C}_2) = V_2$ for fixed V_1, V_2. As shown in [3], Theorem 9.2.7, these are all equal:

$$\max V(\mathcal{C}_1 \sqcup \mathcal{C}_2) = \max V(\mathcal{C}_1 \sqcap \mathcal{C}_2) = \max V(\mathcal{D}_1 \boxtimes \mathcal{D}_2) \equiv S(V_1, V_2).$$

[3] provided the following bound for this common value:

Proposition 1.1. *$S(V_1, V_2) \le T(V_1, V_2)$ where, with ${}_rC_{\le v} \equiv \sum_{j=0}^{v} \binom{r}{j}$,*

$$T(V_1, V_2) \equiv \sup\{r \in \mathbb{N} : \ {}_rC_{\le V_1} \ {}_rC_{\le V_2} \ge 2^r\}. \tag{1.1}$$

Because of the somewhat inexplicit nature of the bound in (1.1), this proposition seems not to have been greatly used so far.

Furthermore, [4] (Theorem 4.27, page 63; Proposition 4.38, page 64) showed that $S(1, k) \le 2k + 1$ for all $k \ge 1$ with equality for $k = 1, 2, 3$.

Here we give a further more explicit bound for $T(V_1, V_2)$ and extend the bounds to the case of general $m \ge 2$. Our main result is the following proposition.

Theorem 1.1. *Let $V \equiv \sum_{j=1}^m V_j$. Then the following bounds hold:*

$$\left\{ \begin{array}{l} V(\sqcup_{j=1}^m \mathcal{C}_j) \\ V(\sqcap_{j=1}^m \mathcal{C}_j) \\ V(\boxtimes_1^m \mathcal{D}_j) \end{array} \right\} \le c_1 V \log\left(\frac{c_2 m}{e^{Ent(\underline{V})/\overline{V}}} \right) \le c_1 V \log(c_2 m), \tag{1.2}$$

where $\underline{V} \equiv (V_1, \ldots, V_m)$, $c_1 \equiv \frac{e}{(e-1)\log(2)} \doteq 2.28231\ldots$, $c_2 \equiv \frac{e}{\log 2} \doteq 3.92165\ldots$,

$$Ent(\underline{V}) \equiv m^{-1} \sum_{j=1}^m V_j \log V_j - \overline{V} \log \overline{V}$$

is the "entropy" of the V_j's under the discrete uniform distribution with weights $1/m$ and $\overline{V} = m^{-1} \sum_{j=1}^m V_j$.

Corollary 1.1. *For $m = 2$ the following bounds hold:*

$$S(V_1, V_2) \le T(V_1, V_2) \le \left\lfloor c_1 (V_1 + V_2) \log\left(\frac{2c_2}{\exp(Ent(\underline{V})/\overline{V})} \right) \right\rfloor \equiv R(V_1, V_2)$$

where c_1, c_2, $Ent(\underline{V})$, and $\overline{V}$ are as in Theorem 1.

Proof. The subsets picked out by $\sqcap_i \mathcal{C}_i$ from a given set of points $\{x_1, \ldots, x_n\}$ in $\mathcal{X}$ are the sets $C_1 \cap \cdots \cap C_m \cap \{x_1, \ldots, x_n\}$. They can be formed by first forming all *different* sets of the form $C_1 \cap \{x_1, \ldots, x_n\}$ for $C_1 \in \mathcal{C}_1$, next intersecting each of these sets by sets in $\mathcal{C}_2$ giving all sets of the form $C_1 \cap C_2 \cap \{x_1, \ldots, x_n\}$, etc. If $\Delta_n(\mathcal{C}, y_1, \ldots, y_n) \equiv \#\{C \cap \{y_1, \ldots, y_n\} : \ C \in \mathcal{C}\}$ and $\Delta_n(\mathcal{C}) = \max_{y_1, \ldots, y_n} \Delta_n(\mathcal{C}, y_1, \ldots, y_n)$ for every collection of sets $\mathcal{C}$ and points $y_1, \ldots, y_n$ (as in [11], page 135), then in the first step we obtain at most $\Delta_n(\mathcal{C}_1)$ different sets, each with n or fewer points. In the second step each of these sets gives rise to at most $\Delta_n(\mathcal{C}_2)$ different sets, etc. We conclude that

$$\Delta_n(\sqcap_i \mathcal{C}_i) \le \prod_i \Delta_n(\mathcal{C}_i) \le \prod_i \left(\frac{en}{V_i}\right)^{V_i},$$

by [11], Corollary 2.6.3, page 136, and the bound $(en/s)^s$ for the number of subsets of size smaller than s for $n \ge s$. By definition the left side of the display is 2^n for n equal to the VC-dimension of $\sqcap_i \mathcal{C}_i$. We conclude that

$$2^n \le \prod_{i=1}^m \left(\frac{en}{V_i}\right)^{V_i},$$

or

$$n \log 2 \le \sum_{i=1}^m V_i \log(e/V_i) + \left(\sum_{i=1}^m V_i\right) \log n.$$

With $V \equiv \sum_i V_i$, define $r = en/V$. Then the last display implies that

$$rV \frac{\log 2}{e} \le \sum_i V_i \log(e/V_i) + V \log(rV/e),$$

or

$$\begin{aligned} r\frac{\log 2}{e} &\le \log r + \log V - \frac{\sum_i V_i \log V_i}{V} \\ &= \log r + \log m - \frac{Ent(\underline{V})}{\overline{V}} = \log\left(\frac{mr}{e^{Ent(\underline{V})/\overline{V}}}\right), \end{aligned}$$

and this inequality can in turn be rewritten as

$$\frac{x}{\log x} \equiv \frac{mr/e^{Ent(\underline{V})/\overline{V}}}{\log\left(mr/e^{Ent(\underline{V})/\overline{V}}\right)} \le \frac{m}{e^{Ent(\underline{V})/\overline{V}}} \cdot \frac{e}{\log 2} \equiv y.$$

Now note that $g(x) \equiv x/\log x \le y$ for $x \ge e$ implies that $x \le (e/(e-1))y \log y$: g is minimized by $x = e$ and is increasing; furthermore $y \ge g(x)$ for $x \ge e$ implies that

$$\log y \ge \log x - \log\log x = \log x \left(1 - \frac{\log\log x}{\log x}\right) \ge \log x \left(1 - \frac{1}{e}\right)$$

so that

$$x \le y \log x \le y \left(1 - \frac{1}{e}\right)^{-1} \log y = \frac{e}{e-1} y \log y.$$

Thus we conclude that

$$\frac{mr}{e^{Ent(\underline{V})/\overline{V}}} \le \frac{e}{e-1} \frac{me}{e^{Ent(\underline{V})/\overline{V}} \log 2} \log\left(\frac{m}{e^{Ent(\underline{V})/\overline{V}}} \cdot \frac{e}{\log 2}\right),$$

which implies that

$$r \le \frac{e^2}{(e-1)\log 2} \log\left(\frac{m}{\exp(Ent(\underline{V})/\overline{V})} \cdot \frac{e}{\log 2}\right).$$

Expressing this in terms of n yields the first inequality (1.2). The second inequality holds since $Ent(\underline{V}) \geq 0$ implies $\exp(Ent(\underline{V})/\overline{V}) \geq 1$.

The corresponding statement for the unions follows because a class $\mathcal{C}$ of sets and the class $\mathcal{C}^c$ of their complements possess the same VC-dimension, and $\cup_i C_i = (\cap_i C_i^c)^c$.

In the case of products, note that

$$\Delta_n(\boxtimes_1^m \mathcal{D}_j) \leq \prod_1^m \Delta_n(\mathcal{D}_j) \leq \prod_{j=1}^m \left(\frac{en}{V_j}\right)^{V_j},$$

and then the rest of the proof proceeds as in the case of intersections. □

It follows from concavity of $x \mapsto \log x$ that with $p_j \equiv V_j / \sum_{i=1}^m V_i$,

$$\frac{\sum_{j=1}^m V_j \log V_j}{\sum_{j=1}^m V_j} = \sum_1^m p_j \log V_j \leq \log\left(\sum_1^m p_j V_j\right) \leq \log\left(\sum_1^m V_j\right)$$

and hence

$$1 \leq \frac{m}{e^{Ent(\underline{V})/\overline{V}}} \leq m, \tag{1.3}$$

or $0 \leq Ent(\underline{V})/\overline{V} \leq \log m$, or

$$0 \leq Ent(\underline{V}) \leq \overline{V} \log m.$$

Here are two examples showing that the quantity $m/e^{Ent(\underline{V})/\overline{V}}$ can be very close to 1 (rather than m) if the V_i's are quite heterogeneous, even if m is large.

Example 1.1. Suppose that $r \in \mathbb{N}$ (large), and that $V_i = r^i$ for $i = 1, \ldots, m$. Then it is not hard to show that

$$\frac{m}{e^{Ent(\underline{V})/\overline{V}}} \to \frac{r}{r-1} r^{1/(r-1)} = \frac{r}{r-1} \exp((r-1)^{-1} \log r)$$

as $m \to \infty$ where the right side can be made arbitrarily close to 1 by choosing r large.

Example 1.2. Suppose that $m = 2$ and that $V_1 = k$, $V_2 = rk$ for some $r \in \mathbb{N}$. Then

$$Ent(\underline{V})/\overline{V} = \log 2 - \frac{1}{r+1} \log((r+1)(1+1/r)^r) \to \log 2$$

as $r \to \infty$ for any fixed k. Therefore

$$\frac{2}{e^{Ent(\underline{V})/\overline{V}}} \to 1$$

as $r \to \infty$ for any fixed k.

Our last example shows that the bound of Theorem 1.1 may improve considerably on the bounds resulting from iteration of Dudley's bound $S(1, k) \leq 2k + 1$.

Example 1.3. Suppose $V_1 = V(\mathcal{C}_1) = k$ and $V_j = V(\mathcal{C}_j) = 1$ for $j = 2, \ldots, m$. Iterative application of Dudley's bound $S(1, k) \leq 2k + 1$ yields $V(\sqcap_{j=1}^m \mathcal{C}_j) \leq 2^{m-1}(k + 1) - 1$, which grows exponentially as $m \to \infty$. On the other hand, Theorem 1.1 yields $V(\sqcap_{j=1}^m \mathcal{C}_j) \leq c_1(m + k - 1) \log(c_2 m)$ which is of order $c_1 m \log m$ as $m \to \infty$.

Although we have succeeded here in providing quantitative bounds for $V(\sqcup_{j=1}^m \mathcal{C}_j)$, $V(\sqcap_{j=1}^m \mathcal{C}_j)$, and $V(\boxtimes_1^m \mathcal{D}_j)$, it seems that we are far from being able to provide quantitative bounds for the VC - dimensions of the (much larger) classes involved in [6], [7], and [10].

Acknowledgement

We owe thanks to a helpful referee for pointing out [4] and for suggesting Example 1.3.

References

[1] Assouad, P. (1983). Densité et dimension. *Ann. Inst. Fourier (Grenoble)* **33** 233–282.
[2] Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929 (1979).
[3] Dudley, R. M. (1984). A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982. Lecture Notes in Math.* **1097** 1–142. Springer, Berlin.
[4] Dudley, R. M. (1999). *Notes on Empirical Processes. MaPhySto Lecture Notes* **4**.
[5] Dudley, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge University Press, Cambridge.
[6] Laskowski, M. C. (1992). Vapnik-Chervonenkis classes of definable sets. *J. London Math. Soc. (2)* **45** 377–384.
[7] Olshen, R. A., Biden, E. N., Wyatt, M. P. and Sutherland, D. H. (1989). Gait analysis and the bootstrap. *Ann. Statist.* **17** 1419–1440.
[8] Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57** 1027–1057.
[9] Pollard, D. (1984). *Convergence of Stochastic Processes. Springer Series in Statistics.* Springer-Verlag, New York.
[10] Stengle, G. and Yukich, J. E. (1989). Some new Vapnik-Chervonenkis classes. *Ann. Statist.* **17** 1441–1446.
[11] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics.* Springer-Verlag, New York. With applications to statistics.