A note on bounds for VC dimensions

Aad van der Vaart and Jon A. Wellner

Abstract: We provide bounds for the VC dimension of class of sets formed by unions, intersections, and products of VC classes of sets $C_1, \ldots, C_m$.

1. Introduction and main results

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

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

then

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

and $V(C) = -1$ if $C$ is empty. (The VC-dimension $V(C)$ defined here corresponds to $S(C)$ as defined by [5] page 134. Dudley, and following him ourselves in [11], used the notation $V(C)$ for the VC-index, which is the dimension plus 1. We have switched to using $V(C)$ 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 $C_1, C_2, \ldots, C_m$ are VC-classes of subsets of a given set $X$ with VC dimensions $V_1, \ldots, V_m$. It is known that the classes $\sqcup_{j=1}^m C_j$, $\sqcap_{j=1}^m C_j$ defined by

$$\sqcup_{j=1}^m C_j \equiv \{ \cup_{j=1}^m C_j : C_j \in C_j, \ j = 1, \ldots, m \},$$

$$\sqcap_{j=1}^m C_j \equiv \{ \cap_{j=1}^m C_j : C_j \in C_j, \ j = 1, \ldots, m \},$$

are again VC: when $C_1 = \cdots = C_m = C$ 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 $C_1$, $C_2$ 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,
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 \( \bigotimes_{j=1}^m \mathcal{D}_j \) defined by
\[
\bigotimes_{j=1}^m \mathcal{D}_j \equiv \{ D_1 \times \cdots \times D_m : D_j \in \mathcal{D}_j, \ j = 1, \ldots, m \}
\]
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 \( \max V(\mathcal{C}_1 \cup \mathcal{C}_2) \), \( \max V(\mathcal{C}_1 \cap \mathcal{C}_2) \), and \( \max V(\mathcal{D}_1 \otimes \mathcal{D}_2) \), 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:
\[
\max V(\mathcal{C}_1 \cup \mathcal{C}_2) = \max V(\mathcal{C}_1 \cap \mathcal{C}_2) = \max V(\mathcal{D}_1 \otimes \mathcal{D}_2) \equiv S(V_1, V_2).
\]

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

**Proposition 1.1.** \( S(V_1, V_2) \leq T(V_1, V_2) \) where, with \( r, \mathcal{C} \leq \nu \equiv \sum_{j=0}^v \binom{\nu}{j} \),

\[
(1.1) \quad T(V_1, V_2) \equiv \sup \{ r \in \mathbb{N} : r, \mathcal{C} \leq V_1, r, \mathcal{C} \leq V_2 \geq 2^r \}.
\]

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 \( S(1, k) \leq 2k + 1 \) for all \( k \geq 1 \) 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 \( m \geq 2 \). Our main result is the following proposition.

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

\[
(1.2) \quad \begin{cases}
V(\bigcup_{j=1}^m \mathcal{C}_j) \\
V(\bigcap_{j=1}^m \mathcal{C}_j) \\
V(\bigotimes_{j=1}^m \mathcal{D}_j)
\end{cases} \leq c_1 V \log \left( \frac{c_2 m}{\bar{\text{Ent}}(V) / \sqrt{V}} \right) \leq c_1 V \log (c_2 m),
\]

where \( \bar{\text{Ent}}(V) \equiv (V_1, \ldots, V_m), c_1 \equiv \frac{e}{(e-1) \log(2)} \equiv 2.28231 \ldots, c_2 \equiv \frac{e}{\log(2)} \equiv 3.92165 \ldots, \)

\[
\bar{\text{Ent}}(V) \equiv m^{-1} \sum_{j=1}^m V_j \log V_j - V \log V
\]
is the “entropy” of the \( V_j \)'s under the discrete uniform distribution with weights \( 1/m \) and \( \bar{V} = m^{-1} \sum_{j=1}^m V_j \).

**Corollary 1.1.** For \( m = 2 \) the following bounds hold:

\[
S(V_1, V_2) \leq T(V_1, V_2) \leq \left[ c_1 (V_1 + V_2) \log \left( \frac{2c_2}{\exp(\bar{\text{Ent}}(V) / \sqrt{V})} \right) \right] \equiv R(V_1, V_2)
\]

where \( c_1, c_2, \bar{\text{Ent}}(V), \) and \( \bar{V} \) are as in Theorem 1.

**Proof.** The subsets picked out by \( \cap_i C_i \) 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 \( \Delta_n(C, y_1, \ldots, y_n) \equiv \# \{ C \cap \{y_1, \ldots, y_n\} : C \in \mathcal{C} \} \) and \( \Delta_n(C) \equiv \max_{y_1, \ldots, y_n} \Delta_n(C, y_1, \ldots, y_n) \) 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 \( \Delta_n(C_1) \) different sets, each with \( n \) or fewer points. In the second step each of these sets gives rise to at most \( \Delta_n(C_2) \) different sets, etc. We conclude that

\[
\Delta_n(\bigcap_i C_i) \leq \prod_i \Delta_n(C_i) \leq \prod_i \left( \frac{en}{V_i} \right)^{V_i},
\]

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 \( n \geq s \). By definition the left side of the display is \( 2^n \) for \( n \) equal to the VC-dimension of \( \bigcap_i C_i \). We conclude that

\[
2^n \leq \prod_{i=1}^m \left( \frac{en}{V_i} \right)^{V_i},
\]

or

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

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

\[
r V \frac{\log 2}{e} \leq \sum_i V_i \log(e/V_i) + V \log(rV/e),
\]

or

\[
r \frac{\log 2}{e} \leq \log r + \log V - \sum_i V_i \log V_i \leq \log r + \log m - \frac{\text{Ent}(V)}{V} = \log \left( \frac{mr}{e^{\text{Ent}(V)/V}} \right),
\]

and this inequality can in turn be rewritten as

\[
x \log x = \frac{mr/e^{\text{Ent}(V)/V}}{\log \left( \frac{mr/e^{\text{Ent}(V)/V}}{e^{\text{Ent}(V)/V}} \right)} \leq \frac{m}{e^{\text{Ent}(V)/V}} \cdot \frac{e}{\log 2} \equiv y.
\]

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

\[
\log y \geq \log x - \log \log x = \log x \left( 1 - \frac{\log \log x}{\log x} \right) \geq \log x \left( 1 - \frac{1}{e} \right)
\]

so that

\[
x \leq y \log x \leq y \left( 1 - \frac{1}{e} \right)^{-1} \log y = \frac{e}{e-1} y \log y.
\]

Thus we conclude that

\[
\frac{mr}{e^{\text{Ent}(V)/V}} \leq \frac{e}{e-1} \cdot \frac{me}{e^{\text{Ent}(V)/V} \log 2} \log \left( \frac{m}{e^{\text{Ent}(V)/V}} \cdot \frac{e}{\log 2} \right),
\]

which implies that

\[
r \leq \frac{e^2}{(e-1) \log 2} \log \left( \frac{m}{\exp(\text{Ent}(V)/V)} \cdot \frac{e}{\log 2} \right).
\]
Expressing this in terms of \( n \) yields the first inequality (1.2). The second inequality holds since \( \text{Ent}(V) \geq 0 \) implies \( \exp(\text{Ent}(V)/V) \geq 1 \).

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
\[
\Delta_n(\mathbb{E}_{i=1}^m D_j) \leq \prod_{i=1}^m \Delta_n(D_j) \leq \prod_{j=1}^m \left( \frac{e V}{V_j} \right)^{V_j},
\]
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 \( p_j \equiv V_j/\sum_{i=1}^m V_i \),
\[
\sum_{j=1}^m \frac{V_j \log V_j}{\sum_{j=1}^m V_j} = \sum_{j=1}^m p_j \log V_j \leq \log \left( \sum_{j=1}^m p_j V_j \right) \leq \log \left( \sum_{j=1}^m V_j \right)
\]
and hence
\[
1 \leq \frac{m}{e^{\text{Ent}(V)/V}} \leq m,
\]
or \( 0 \leq \text{Ent}(V)/V \leq \log m \), or
\[
0 \leq \text{Ent}(V) \leq V \log m.
\]

Here are two examples showing that the quantity \( m/e^{\text{Ent}(V)/V} \) 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
\[
\frac{m}{e^{\text{Ent}(V)/V}} \rightarrow \frac{r}{r - 1} r^{1/(r-1)} = \frac{r}{r - 1} \exp((r - 1)^{-1} \log r)
\]
as \( m \rightarrow \infty \) 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
\[
\text{Ent}(V)/V = \log 2 - \frac{1}{r + 1} \log((r + 1)(1 + 1/r)^r) \rightarrow \log 2
\]
as \( r \rightarrow \infty \) for any fixed \( k \). Therefore
\[
\frac{2}{e^{\text{Ent}(V)/V}} \rightarrow 1
\]
as \( r \rightarrow \infty \) 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 \( V_1 = V(C_1) = k \) and \( V_j = V(C_j) = 1 \) for \( j = 2, \ldots, m \). Iterative application of Dudley’s bound \( S(1, k) \leq 2k + 1 \) yields \( V(\cap_{j=1}^m C_j) \leq 2^{m-1}(k + 1) - 1 \), which grows exponentially as \( m \rightarrow \infty \). On the other hand, Theorem 1.1 yields \( V(\cap_{j=1}^m C_j) \leq c_1 (m + k - 1) \log(c_2 m) \) which is of order \( c_1 m \log m \) as \( m \rightarrow \infty \).
Although we have succeeded here in providing quantitative bounds for $V(\bigcup_{j=1}^{m}C_j)$, $V(\bigcap_{j=1}^{m}C_j)$, and $V(\bigoplus_{j=1}^{m}D_j)$, 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