

# VC Dimension: Examples and Tools

Carlos C. Rodríguez  
*<http://omega.albany.edu:8008/>*

October 5, 2004

## Recall the Definitions

Let  $\mathcal{A}$  be a class of subsets of  $R^d$ . The shatter coefficients of  $\mathcal{A}$  are denoted by  $\mathcal{S}(n, \mathcal{A})$  and defined as the maximum number of subsets of a set of  $n$  elements that appear in the elements of  $\mathcal{A}$ . We say that the class  $\mathcal{A}$  shatters a given set of  $n$  elements if each of its  $2^n$  subsets are picked by the members of  $\mathcal{A}$ . The Vapnik-Chervonenkis (VC) dimension of  $\mathcal{A}$  is the size of the largest set that can be shattered by  $\mathcal{A}$ . It is denoted by  $V = V(\mathcal{A})$ . Thus, if  $V < \infty$  we must have  $\mathcal{S}(n, \mathcal{A}) = 2^n$  for each  $n \leq V$  and  $\mathcal{S}(V+1, \mathcal{A}) < 2^{V+1}$ . In words: If the VC dimension is  $V$  the class  $\mathcal{A}$  shatters some sets with  $V$  elements but it shatters no set with  $V+1$  (or more) elements. Here are the simplest examples.

## Simplest Examples

**Half lines in  $R$**   $\mathcal{A} = \{(-\infty, x] : x \in R\}$ . Clearly,  $\mathcal{S}(n, \mathcal{A}) = n+1$  and so  $V = 1$ . Notice, that,

$$\mathcal{S}(n, \mathcal{A}) = n+1 = \binom{n}{0} + \binom{n}{1}$$

**Intervals in  $R$**   $\mathcal{A} = \{(a, b) : a, b \in R\}$ . To compute  $\mathcal{S}(n, \mathcal{A})$  just think of the  $(n+1)$  spaces defined by  $n$  different real numbers. Any choice of two of these spaces corresponds to an  $(a, b)$  picking up the points inside. Thus,

$$\mathcal{S}(n, \mathcal{A}) = \binom{n+1}{2} + 1$$

where the extra interval is the one picking none of the  $n$  points. Therefore,  $V = 2$  and again,

$$\mathcal{S}(n, \mathcal{A}) = \frac{n(n+1)}{2} + 1 = \binom{n}{0} + \binom{n}{1} + \binom{n}{2}$$

**Convex Polygons in  $R^2$**  It is easy to see that for this class  $V = \infty$ . Just think of the  $n$  points lying on the unit circle. Every subset of these  $n$  points is picked by the convex polygon having the elements of the subset as vertices.

To be able to build new examples easily, we show some general properties of shatter coefficients.

Let  $\mathcal{A}$  and  $\mathcal{B}$  be two classes of subsets of  $R^d$  and let  $*$  be a set operation. We define,

$$\mathcal{A} (*) \mathcal{B} = \{A * B : A \in \mathcal{A}, B \in \mathcal{B}\}$$

We have,

1.  $\mathcal{S}(n, \mathcal{A}^{(c)}) = \mathcal{S}(n, \mathcal{A})$ .
2.  $\mathcal{S}(n, \mathcal{A} (\cap) \mathcal{B}) \leq \mathcal{S}(n, \mathcal{A}) \mathcal{S}(n, \mathcal{B})$ .
3.  $\mathcal{S}(n, \mathcal{A} (\cup) \mathcal{B}) \leq \mathcal{S}(n, \mathcal{A}) \mathcal{S}(n, \mathcal{B})$ .
4.  $\mathcal{S}(m+n, \mathcal{A}) \leq \mathcal{S}(m, \mathcal{A}) \mathcal{S}(n, \mathcal{A})$ .
5.  $\mathcal{S}(n, \mathcal{A} (\times) \mathcal{B}) \leq \mathcal{S}(n, \mathcal{A}) \mathcal{S}(n, \mathcal{B})$ .

**Proof:** The first one is trivial. Let's show the second. Consider a fix set  $\{x_1, \dots, x_n\}$  of  $n$  points. The class  $\mathcal{A}$  picks the subsets  $A_1, A_2, \dots, A_k$  with  $k \leq \mathcal{S}(n, \mathcal{A})$ . Now if  $A_j$  has  $n_j$  elements then  $\mathcal{B}$  picks at most  $\mathcal{S}(n_j, \mathcal{B})$  subsets of  $A_j$ . Thus,

$$\mathcal{S}(n, \mathcal{A} (\cap) \mathcal{B}) \leq \sum_{j=1}^{\mathcal{S}(n, \mathcal{A})} \mathcal{S}(n_j, \mathcal{B})$$

the result follows by noticing that  $n_j \leq n$ . Three follows from one and two. To show four proceed as above by choosing a set of  $(m+n)$  points with  $m$  of the points marked and defining  $A_1, \dots, A_k$  now with  $k = \mathcal{S}(m+n, \mathcal{A})$  and split each  $A_j$  as,

$$A_j = A_{j1} \cup A_{j2}$$

with  $A_{j1}$  containing all of the marked points in  $A_j$ . The total number of sets  $A_{j1}$  is at most  $\mathcal{S}(m, \mathcal{A})$  and the total number of different sets of type  $A_{j2}$  is at most  $\mathcal{S}(n, \mathcal{A})$ . Thus,  $k \leq \mathcal{S}(m, \mathcal{A}) \mathcal{S}(n, \mathcal{A})$ . The proof for five is similar to the second case. •

## More Examples

It is not difficult to generalize the previous examples to  $R^d$ .

**South-West Intervals in  $R^d$**   $\mathcal{A} = \{(-\infty, a_1] \times (-\infty, a_2] \dots \times (-\infty, a_d] : a_j \in R\}$ . By what we found for the case  $d = 1$  and property 5 above, we can deduce that,  $\mathcal{S}(n, \mathcal{A}) \leq (n + 1)^d$  and looking ahead we guess  $V = d$ . This can be shown by noticing that  $\mathcal{A}$  shatters the  $d$  canonical vectors  $E = \{e_1, \dots, e_d\}$  with  $e_j = (0, \dots, 0, 1, 0, \dots, 0)$  (i.e. all zeroes except a 1 in the  $j$ th entry). Any  $B \subset E$  is picked by a member of  $\mathcal{A}$ , e.g., by the one with  $a_j = 1 + \epsilon_j/2$  where  $\epsilon_j = +1$  if  $e_j \in B$  and  $\epsilon_j = -1$  otherwise. Furthermore, no set with  $d + 1$  points can be shattered by  $\mathcal{A}$  since it is impossible to pick only the  $d$  points where the first has the largest first coordinate, the second has the largest second coordinate, ..., the  $d$ th has the largest  $d$ th coordinate, and not the other. Thus,  $V = d$ . •

**All the rectangles in  $R^d$**  For this class  $V = 2d$ . Again the case  $d = 1$  and property 5 above shows that,

$$\mathcal{S}(n, \mathcal{A}) \leq \left( \frac{n(n+1)}{2} + 1 \right)^d < (n+1)^{2d}$$

and looking ahead we'd guess  $V = 2d$ . A proof along the lines of the previous example is straight forward.

## Power Tools

The following simple result is known as Sauer's Lemma. It provides a way to bound shatter coefficients in terms of the VC dimension.

**Sauer's Lemma** If  $V(\mathcal{A}) < \infty$ , for all  $n$

$$\mathcal{S}(n, \mathcal{A}) \leq \sum_{i=0}^V \binom{n}{i}$$

**Proof:** By induction on  $n$ . It is true for  $n \leq V$  since for these  $n$ ,

$$\mathcal{S}(n, \mathcal{A}) = 2^n = \sum_{i=0}^n \binom{n}{i} = \sum_{i=0}^V \binom{n}{i}.$$

Now let us assume the result true for  $n$  and deduce it for  $n + 1$ . Clearly  $\mathcal{S}(n + 1, \mathcal{A})$  is at most  $2\mathcal{S}(n, \mathcal{A})$  (see property 4 above) but this bound can be reduced by using the fact that the vc dimension is  $V$  so we know that  $\mathcal{A}$  shatters no set with  $V + 1$  points. On the other hand, a set with  $n + 1$  points has  $n$  choose  $V$  more subsets with  $V + 1$  elements than a set with  $n$  points has, and each of these must hide at least one new subset that cannot be picked out by  $\mathcal{A}$ .

Hence, applying the induction hypothesis, and the formula for Pascal's triangle satisfied by the binomial coefficients, we obtain,

$$\begin{aligned}
\mathcal{S}(n+1, \mathcal{A}) &\leq 2\mathcal{S}(n, \mathcal{A}) - \binom{n}{V} \\
&= 2 \sum_{i=0}^V \binom{n}{i} - \binom{n}{V} \\
&\leq \sum_{i=0}^V \binom{n}{i} + \sum_{i=1}^V \binom{n}{i-1} \\
&= \sum_{i=0}^V \binom{n+1}{i}.
\end{aligned}$$

The promised bounds follow.

**Corollary** If  $V < \infty$ , for all  $n$ ,

$$\mathcal{S}(n, \mathcal{A}) \leq (n+1)^V,$$

and when  $n \geq V$ ,

$$\mathcal{S}(n, \mathcal{A}) \leq \left(\frac{ne}{V}\right)^V.$$

**Proof:** We have,

$$\sum_{i=0}^V \binom{n}{i} \leq \sum_{i=0}^V \frac{n^i}{i!} \leq \sum_{i=0}^V \frac{n^i V!}{i!(V-i)!} = (n+1)^V$$

Also, if  $V/n \leq 1$ ,

$$\begin{aligned}
\left(\frac{V}{n}\right)^V \sum_{i=0}^V \binom{n}{i} &\leq \sum_{i=0}^V \left(\frac{V}{n}\right)^i \binom{n}{i} \\
&\leq \sum_{i=0}^n \left(\frac{V}{n}\right)^i \binom{n}{i} \\
&= \left(1 + \frac{V}{n}\right)^n \leq e^V.
\end{aligned}$$

The other useful tool is,

**Theorem** If  $\mathcal{G}$  is an  $m$ -dimensional vector space of real-valued functions defined on  $R^d$  then the class of sets,

$$\mathcal{A} = \{ \{x : g(x) \geq 0\} : g \in \mathcal{G} \}$$

has VC dimension  $V \leq m$ .

**Proof:** Take a set of  $m+1$  arbitrary points  $\{x_1, \dots, x_{m+1}\}$  in  $R^d$ . We show one subset that can never be picked out by  $\mathcal{A}$ . To define this special subset consider the linear transformation  $L$  from  $\mathcal{G}$  to  $R^{m+1}$  defined by,

$$L(g) = (g(x_1), \dots, g(x_{m+1})).$$

Since the image space  $L\mathcal{G}$  is of dimension at most  $m$ , there must exist a non-zero vector  $\gamma \in R^{m+1}$  orthogonal to the image space  $L\mathcal{G}$ . Thus, there exist scalars  $\gamma_1, \dots, \gamma_{m+1}$  not all zero such that,

$$\sum_{i=1}^{m+1} \gamma_i g(x_i) = 0 \text{ for each } g \in \mathcal{G}.$$

Separate the sum above according to whether  $\gamma_i$  is negative ( $i \in \{-\}$ ) or not ( $i \in \{+\}$ ), to obtain,

$$\sum_{\{+\}} \gamma_i g(x_i) = \sum_{\{-\}} (-\gamma_i) g(x_i).$$

By replacing  $\gamma$  by  $-\gamma$  if necessary, we may assume (wlog since  $\gamma \neq 0$ ) that  $\{-\}$  is non-empty and this shows that it is impossible for  $\mathcal{A}$  to only pick out the  $x_i$  with  $i \in \{+\}$ . For if there was a  $g \in \mathcal{G}$  with  $\{x : g(x) \geq 0\}$  performing the trick, it would have to make the lhs of the above equation  $\geq 0$  but the rhs  $< 0$  since  $\{-\}$  is non-empty and  $g(x_i) < 0$  for all  $i \in \{-\}$ . •