Properties of the V-C Dimension

Andrew Fyfe

Computer Science Department
California Institute of Technology

Caltech-CS-TR-90-01

Properties of the V-C Dimension

Thesis by
Andrew Fyfe

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science

CALTECH-CS-TR-90-01

California Institute of Technology
Pasadena, California

1989 November 21

Contents

List of Figures

Chapter 1

Introduction

Consider a system in which we teach a neural network a function f by providing example inputs and the corresponding binary output. For each possible function g that the network is capable of evaluating, there will a corresponding probability, π_g, that $f = g$. Based on the examples provided, the observed frequency of success, ν_g, is equal to the number of examples where $f = g$, divided by the total number of examples presented.

If we can assert that, with high probability, π_g and ν_g do not differ by more than ϵ, then we say that the neural network *generalizes*[AM89]. If we have a single function g that the network can compute, we can use Bernoulli's Theorem to demonstrate generalization, for, after N examples,

$$Pr[\,|\nu - \pi| > \epsilon\,] \leq \frac{1}{4N\epsilon^2}.$$

In [VC71], Vapnik and Chervonenkis generalized this result, for $N \geq 2/\epsilon^2$, to

$$Pr[\sup_g |\nu_g - \pi_g| > \epsilon\,] \leq 4 \times m(2N) \times e^{-\epsilon^2 N/8}.$$

Vapnik and Chervonenkis further showed that the function $m(n)$ is either identically equal to 2^n or is bounded by $n^{d+1} + 1$, depending on the structure of the set of functions $\{g\}$. Should $m(n)$ be bounded by a polynomial, then we have uniform convergence of the observed frequencies to the actual probabilities, and the neural network will generalize. We need to have the set $\{g\}$ be infinite, for $m(n) = 2^n$.

The constant d is refered to as the V-C dimension. We say $d = \infty$ if $m(n) = 2^n$.

Just as the Kolmogorov complexity, or entropy, of a binary string gives us some information about the string, the V-C dimension of a set of functions likewise gives us some information. In order to pursue this further, we will first investigate some basic properties of the V-C dimension. In particular, we will develop an upper and lower bound on d, given only the size of the set of functions $\{g\}$.

Chapter 2

Basic Definitions

Let $\mathcal{X}$ be a finite set, $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, $|\mathcal{X}| = n$. We will be considering subsets $F \subseteq \mathcal{F} = \{f : \mathcal{X} \to \{0,1\}\}$, with $|F| = m$, $1 \leq m \leq |\mathcal{F}| = 2^n$.

It is convenient to think of functions in terms of their characteristic strings. For $f \in \mathcal{F}$, we equate f with the binary string $\chi_f = \nu_1\nu_2\ldots\nu_n \in \{0,1\}^n$, where $\nu_i = f(x_i)$. The set $\mathcal{F}$ would be equated with all n-bit binary strings $\{0,1\}^n$.

Suppose $X \subseteq \mathcal{X}$, $X \neq \emptyset$. The function f_X will be the restriction of f to X, defined by $f_X : X \to \{0,1\}$ with $f_X(x) = f(x)$, $\forall x \in X$. $\mathcal{F}_X$ will be the set of all such functions, namely $\mathcal{F}_X = \{\, f_X \mid f \in \mathcal{F} \,\}$, $|\mathcal{F}_X| = 2^{|X|}$.

Definition 2.1 Suppose $X \subseteq \mathcal{X}$. We say that $X \neq \emptyset$ is *shattered* by F if $\{\, f_X \mid f \in F \,\} = \mathcal{F}_X$. This means that given any assignment of $\{0,1\}$ to the elements of X, there is some $f \in F$ with $f(x)$ precisely that same assignment. If $X = \emptyset$, then X is shattered by F if $F \neq \emptyset$.

Definition 2.2 Suppose $\exists X \subseteq \mathcal{X}$ with $|X| = r$ and X is shattered by F, and, $\forall X \subseteq \mathcal{X}$ with $|X| = r+1$, X is not shattered by F. Then we say that the V-C dimension $VC(F) = r$. Notice that if $X = \mathcal{X}$, and X is shattered by F, then $VC(F) = n$. As there can be no subset $X \subseteq \mathcal{X}$ with $|X| = n+1$, the second condition for n to be the V-C dimension is satisfied vacuously. Also, as $F \neq \emptyset$, we always have $X = \emptyset$ shattered by F, and $VC(F) \geq 0$.

Example: If $VC(F) = r$, then $|F| \geq 2^r$, since $\exists X \subseteq \mathcal{X}$ with $|X| = r$ which is shattered by F. This requires at least 2^r distinct functions from F.

Example: $VC(\mathcal{F}) = n$, since $\mathcal{X}$ is shattered by $\mathcal{F}$. Also, if $VC(F) = n$, then $F = \mathcal{F}$.

Example: $VC(\{f\}) = 0$, since there must be at least two functions to shatter even a single element $x \in \mathcal{X}$.

Example: If $f \neq g$, then $VC(\{f, g\}) = 1$, since there must be some element $x \in \mathcal{X}$ with $f(x) \neq g(x)$. Then $X = \{x\}$ is shattered by $\{f, g\}$.

Example: If $F \subseteq G$, then $VC(F) \leq VC(G)$, for if any subset $X \subseteq \mathcal{X}$ is shattered by F, it must also be shattered by G.

Chapter 3

Bounds on the V-C Dimension

Suppose we have some set, F, and we know that $VC(F) = r$. What is the smallest such set, and what is the largest? Given this information, we can transform it into formulas for the upper and lower bounds on the V-C dimension of F, given its size, $m = |F|$.

3.1 Smallest F with $VC(F) = r$

As was observed in an earlier example, we must have $|F| \geq 2^r$ if $VC(F) = r$, as we need at least 2^r functions to shatter an r–subset of $\mathcal{X}$. We will construct a set, F, with $|F| = 2^r$ and $VC(F) = r$. Such a set will be a smallest $F \subseteq \mathcal{F}$ with $VC(F) = r$.

Let $X = \{x_1, x_2, \ldots, x_r\}$, and consider $\mathcal{F}_X$, the set of all boolean functions on X. Here we have $|\mathcal{F}_X| = 2^r$. Now, for each $f \in \mathcal{F}_X$, we define an extension $f' : \mathcal{X} \to \{0, 1\}$ by

$$f'(x) = \begin{cases} f(x) & \text{if } x \in X, \\ 0 & \text{otherwise.} \end{cases}$$

The function f' is defined so that $f'_X = f$. The actual way that f' is extended from X to $\mathcal{X}$ is not important. Let $F = \{\, f' \mid f \in \mathcal{F}_X \,\}$. Then we have $|F| = 2^r$ and $VC(F) \geq r$, as the subset X shatters F. Furthermore, we must have $VC(F) < r + 1$ as $|F| < 2^{r+1}$, hence, $VC(F) = r$, as required.

3.2 Largest F with $VC(F) = r$

As in the previous section, we want to construct a set, F, with $VC(F) = r$. This time, however, we want to show that if we have any other set, G, with $|G| > |F|$, then $VC(G) > VC(F)$. Such an F then will be a largest $F \subseteq \mathcal{F}$ with $VC(F) = r$. The following theorem provides exactly such an F.

Theorem 3.1

1. $\exists F$ with $|F| = \sum_{i=0}^{r} \binom{n}{i}$ and $VC(F) = r$,

2. if $|F| > \sum_{i=0}^{r} \binom{n}{i}$ then $VC(F) > r$.

First, we need a lemma.

Lemma 3.2 Let $\mathcal{B} = \{0,1\}^n$, the set of all n–bit binary strings. We will define subsets $B \subseteq \mathcal{B}$, called *volumes*, about a selection of r of the n bits. The volume B will be the all those strings that assign a given pattern of ones and zeros to the selected r bits, leaving the remaining $n - r$ bits free to take on all of the 2^{n-r} possible assignments. Hence $|B| = 2^{n-r}$.

Suppose we select volumes $B_1, B_2, \ldots, B_{\binom{n}{r}}$, each representing one of the $\binom{n}{r}$ possible selections of r bits, and consider

$$C = \bigcup_{i=1}^{\binom{n}{r}} B_i.$$

Then $|C| \geq \sum_{i=r}^{n} \binom{n}{i}$.

Before we prove the lemma, let us consider the following example in detail.
Example: Suppose $n = 4$. We select bits 2 and 3 ($r = 2$) and give the assignment 0,1. Then the volume $B = \{0010, 0011, 1010, 1011\}$. We will abbreviate this B by $*01*$.

Suppose we take the following example: $n = 4$, $r = 2$, and $B_1 = 11{*}{*}$, $B_2 = 0{*}1{*}$, $B_3 = 0{*}{*}0$, $B_4 = {*}11{*}$, $B_5 = {*}0{*}0$, $B_6 = {*}{*}01$. Then $C = \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1100, 1101, 1110, 1111\}$, $|C| = 15$.

We will generate a modified set of volumes B_i' by changing those volumes whose first bit is 0 to have the first bit 1. Hence we get $B_1' = 11{*}{*}$, $B_2' = 1{*}1{*}$, $B_3' = 1{*}{*}0$, $B_4' = {*}11{*}$, $B_5' = {*}0{*}0$, $B_6' = {*}{*}01$, with $C' = \{0000, 0001, 0010, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111\}$, $|C'| = 14$. Note that except for B_2' and B_3', $B_i' = B_i$.

We classify the volumes B_i' based on the first bit of the corresponding B_i as follows: $\mathcal{V}_0 = \{B_2', B_3'\}$, $\mathcal{V}_1 = \{B_1'\}$, $\mathcal{V}_* = \{B_4', B_5', B_6'\}$. We now define three subsets of C' by taking the union of the volumes in each of the three sets above, namely, $V_0 = B_2' \cup B_3' = \{1000, 1010, 1011, 1100, 1110, 1111\}$, $V_1 = B_1' = \{1100, 1101, 1110, 1111\}$, $V_* = B_4' \cup B_5' \cup B_6' = \{0000, 0001, 0010, 0101, 0110, 0111, 1000, 1001, 1010, 1101, 1110, 1111\}$. Note that $C' = V_0 \cup V_1 \cup V_*$.

We will now show that $|C'| \leq |C|$ by showing that for every $\boldsymbol{b} \in C'$ there is a corresponding element, or pair of elements, in C.

Take some binary string $\boldsymbol{b} = b_1b_2b_3b_4 \in C'$. Suppose $\boldsymbol{b} \in V_*$. Then both $0b_2b_3b_4$ and $1b_2b_3b_4$ are in V_* and, hence, C', as the volumes making up V_* leave the first bit free; furthermore, both are in C, as those volumes are the same as the corresponding volumes for C.

There are a number of such elements in $V_* \subseteq C'$; for example, $\boldsymbol{b} = 0010$. We have both 0010 and 1010 in C' and in C.

Suppose $\boldsymbol{b} \notin V_*$. Then $\boldsymbol{b} = 1b_2b_3b_4$, as all the volumes in $\mathcal{V}_0 \cup \mathcal{V}_1$ specify the first bit to be 1, and $0b_2b_3b_4 \notin C'$, as we would have to have $0b_2b_3b_4 \in V_*$, as only V_* contains elements that start with 0, and hence $\boldsymbol{b} = 1b_2b_3b_4 \in V_*$, as V_* contains only pairs $0b_2b_3b_4$ and $1b_2b_3b_4$. If we have $\boldsymbol{b} \in V_0$ and $\boldsymbol{b} \notin V_1$, then we have $0b_2b_3b_4 \in C$ and $1b_2b_3b_4 \notin C$. If instead we have $\boldsymbol{b} \notin V_0$ and $\boldsymbol{b} \in V_1$, then we have $0b_2b_3b_4 \notin C$ and $1b_2b_3b_4 \in C$. Finally, if $\boldsymbol{b} \in V_0 \cap V_1$, then we have both $0b_2b_3b_4 \in C$ and $1b_2b_3b_4 \in C$.

In this example there are only two strings that are in $V_0 \cup V_1$ and not in V_*. They are $\boldsymbol{b}_1 = 1011$ and $\boldsymbol{b}_2 = 1100$. The first, $\boldsymbol{b}_1$, is only in V_0; hence the string 0011 appears in C, while $\boldsymbol{b}_1$ itself does not. The other, $\boldsymbol{b}_2$ is in both V_0 and V_1; hence, both 0100 and 1100 are in C, while only one, 1100, is in C'.

We've shown a correspondence between elements of C' and elements of C. This correspondence may be one for one, or there may be one element in C' for two in C. In the above example, we had just one case where an element of C' was associated with a pair from C (namely $\boldsymbol{b} = 1100$), and we have $|C| = |C'| + 1$. Note that the choices above, namely changing the first bit of the volumes rather than any other, and changing that bit to 1 rather than to 0, are both arbitrary.

Proof of Lemma 3.2: First we select a single bit position. Since we can always reorder the bits, we can assume, without loss of generality, that we selected the first bit.

For each volume B_i, we form a new volume, B_i', as follows: If the first bit of B_i was assigned either zero or one, the first bit of B_i' is assigned a one, and all other bits in B_i' are assigned as in B_i. From these new volumes we get

$$C' = \bigcup_{i=1}^{\binom{n}{r}} B_i'.$$

We will show $|C'| \leq |C|$.

The volumes B_i' can be grouped into three categories, $\mathcal{V}_0$, $\mathcal{V}_1$, and $\mathcal{V}_*$, depending on whether the first bit has been set to 0 or 1, or has been left unspecified in the corresponding B_i. Note that only the $B_i' \in \mathcal{V}_0$ are different from the corresponding B_i, and then only in the first bit.

Let $V_0 = \bigcup_{B_i' \in \mathcal{V}_0} B_i'$, $V_1 = \bigcup_{B_i' \in \mathcal{V}_1} B_i'$, and $V_* = \bigcup_{B_i' \in \mathcal{V}_*} B_i'$. Note that $C' = V_0 \cup V_1 \cup V_*$.

Let $\boldsymbol{c} \in C'$, $\boldsymbol{c} = c_1 c_2 \ldots c_n$. Suppose $\boldsymbol{c} \in V_*$. Then both $0c_2 \ldots c_n$ and $1c_2 \ldots c_n$ are in C', and also in C.

Instead, suppose $\boldsymbol{c} \notin V_*$. Then $\boldsymbol{c} = 1c_2 \ldots c_n$, and $0c_2 \ldots c_n \notin C'$. Since $\boldsymbol{c} \in V_0 \cup V_1$, either $0c_2 \ldots c_n \in C$, if $\boldsymbol{c} \in V_0$ or $1c_2 \ldots c_n \in C$, if $\boldsymbol{c} \in V_1$, or both, if $\boldsymbol{c} \in V_0 \cap V_1$.

Hence, $|C| \geq |C'|$.

By repeating the above for each bit position, we get a set of volumes all of whose specified bits are 1. Each volume specifies a set of strings all of which have at least r ones. The corresponding union C attains the minimum size. In this case $C = \{\, \boldsymbol{b} \in \mathcal{B} \mid \boldsymbol{b}$ contains at least r ones $\}$, with $|C| = \sum_{i=r}^{n} \binom{n}{i}$.

Thus $\forall C, |C| \geq \sum_{i=r}^{n} \binom{n}{i}$. ∎

We showed in the lemma that at each step of the transformation we have $|C| \geq |C'|$. We get equality if $(V_0 \cap V_1) \setminus V_* = \emptyset$. In particular, should $\mathcal{V}_1 = \emptyset$, which would mean that all the volumes B_i either had the first bit set to 0 or left unspecified, we would have $|C| = |C'|$.

Proof of Theorem 3.1: Consider the set of functions, each of whose characteristic string contains at most r ones, namely $F_1 = \{ f \in \mathcal{F} \mid \sum_{x \in \mathcal{X}} f(x) \leq r \}$. Since there are n positions in the characteristic string of a function, of which at most r contain a 1, the number of such functions is $|F_1| = \sum_{i=0}^{r} \binom{n}{i}$.

Claim: $VC(F_1) = r$. First, if $r = n$, then $F_1 = \mathcal{F}$, and $VC(F_1) = VC(\mathcal{F}) = n$. Suppose $r < n$. Let $X = \{x_1, x_2, \ldots, x_r\}$. Then X is shattered by F_1, since given any assignment of 0 or 1 to each $x \in X$, together with the assignment of 0 to $x \notin X$, we obtain a function f that contains at most r ones in its characteristic string, hence, $f \in F_1$.

Furthermore, $VC(F_1) < r + 1$, since no subset $X \subseteq \mathcal{X}$ with $|X| = r + 1$ can shatter F_1, as no $f \in F_1$ can assign all ones to those $x \in X$. Hence, we have $VC(F_1) = r$, as required.

Suppose we have some $F \subseteq \mathcal{F}$, with $VC(F) \leq r$. Let $X_1, X_2, \ldots, X_{\binom{n}{r+1}}$ be the $\binom{n}{r+1}$ $(r+1)$–subsets of $\mathcal{X}$. Then, for any X_i, $|X_i| = r + 1$, there is one assignment, and possibly several assignments, of zeros and ones to each $x \in X_i$ that is avoided by the functions $f \in F$.

For each subset X_i, let $(f_i)_{X_i}$ be a function such that $\forall f \in F$, $f_{X_i} \neq (f_i)_{X_i}$. Such a function, $(f_i)_{X_i}$, exists, as there is some assignment to the $r + 1$ elements of X_i that is missed by all $f \in F$, and this defines such a function, $(f_i)_{X_i}$. Now, define an associated volume $B_i = \left\{ f \in \mathcal{F} \,\middle|\, f_{X_i} = (f_i)_{X_i} \right\}$, the set of all extensions of f_{X_i} to $\mathcal{X}$, and hence $|B_i| = 2^{n-(r+1)}$. In terms of characteristic strings, these volumes B_i are identical to the volumes defined in Lemma 3.2, though on $(r+1)$–subsets, rather than r–subsets.

We will have $\binom{n}{r+1}$ such volumes, $B_1, B_2, \ldots, B_{\binom{n}{r+1}}$, one for each subset X_i. For each B_i, $B_i \cap F = \emptyset$, because $\forall f \in B_i$, $f_{X_i} = (f_i)_{X_i}$ and $\forall f \in F$, $f_{X_i} \neq (f_i)_{X_i}$. If $F^c = \mathcal{F} \setminus F$, then

$$\bigcup_{i=1}^{\binom{n}{r+1}} B_i \subseteq F^c.$$

This gives us

$$\left| \bigcup_{i=1}^{\binom{n}{r+1}} B_i \right| \leq |F^c| = 2^n - |F|,$$

or

$$|F| \leq 2^n - \left| \bigcup_{i=1}^{\binom{n}{r+1}} B_i \right|.$$

This collection of $\binom{n}{r+1}$ volumes corresponds exactly to the collection of volumes of Lemma 3.2, if we think of the binary strings in the lemma as characteristic strings for the functions here. From that lemma we get

$$\left| \bigcup_{i=1}^{\binom{n}{r+1}} B_i \right| \geq \sum_{i=r+1}^{n} \binom{n}{i},$$

and, thus,

$$\begin{aligned} |F| &\leq 2^n - \left| \bigcup_{i=1}^{\binom{n}{r+1}} B_i \right| \\ &\leq 2^n - \sum_{i=r+1}^{n} \binom{n}{i} \\ &= \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=r+1}^{n} \binom{n}{i} \\ &= \sum_{i=0}^{r} \binom{n}{i}. \end{aligned}$$

Hence, if $|F| > \sum_{i=0}^{r} \binom{n}{i}$, then $VC(F) > r$. ∎

3.3 Bounds on the V-C Dimension

We have found, for a given V-C dimension r, a largest and a smallest set F with $VC(F) = r$. We can use this information to provide a formula for the upper and lower bounds on the V-C dimension for an arbitrary set F given its size $m = |F|$, as in the following corollary.

Corollary 3.3 Let $F \subseteq \mathcal{F}$, with $|F| = m$. Then,

$$\min VC(m) = \min \left\{ r \,\middle|\, \sum_{i=0}^{r} \binom{n}{i} \geq m \right\}, \quad (3.1)$$

$$\max VC(m) = \lfloor \log_2 m \rfloor. \quad (3.2)$$

Proof: If $m = 1$, both equations give 0; if $m = 2^n$, both equations give n, as required. So, suppose otherwise, that $1 < m < 2^n$.

For equation 3.1, first recall that $\exists F$ with $|F| = \sum_{i=0}^{r} \binom{n}{i}$ and $VC(F) = r$. Here take $r > 0$, as otherwise we would have $m = 1$.

Consider any subset $G \subseteq F$. We must have $VC(G) \leq VC(F) = r$. However, if $|G| > \sum_{i=0}^{r-1} \binom{n}{i}$, we have that $VC(G) > r - 1$. Thus, for $\sum_{i=0}^{r-1} \binom{n}{i} < m \leq \sum_{i=0}^{r} \binom{n}{i}$, we have that the minimum V-C dimension must be both $> r - 1$ and $\leq r$. Hence, $\min VC(m) = r$ for $\sum_{i=0}^{r-1} \binom{n}{i} < m \leq \sum_{i=0}^{r} \binom{n}{i}$, and equation 3.1 is verified.

For equation 3.2, we first construct a set F, with $|F| = 2^r$ and $VC(F) = r$. In this case, take $r < n$, as otherwise we would have $m = 2^n$. We do this by taking all possible functions $f : X \to \{0, 1\}$ with $|X| = r$. We then extend each f to be a function on $\mathcal{X}$ by choosing any arbitrary assignment for the remaining $x \in \mathcal{X} \setminus X$. These functions make up the required F.

Consider any superset G, $F \supseteq G \supseteq \mathcal{F}$. We must have $VC(G) \geq VC(F) = r$. However, if $|G| < 2^{r+1}$, then $VC(G) < r + 1$. Thus, for $2^r \leq m < 2^{r+1}$, we have that the maximum V-C dimension must be both $< r + 1$ and $\geq r$. Hence, $\max VC(m) = r$ for $2^r \leq m < 2^{r+1}$, and equation 3.2 is verified. ∎

The functions $\min VC(m)$ and $\max VC(m)$ are plotted in Figures 3.1 and 3.2, for the case $n = 48$. In the former, the m axis is linear, while in the latter is it logarithmic. In the case of $\max VC(m)$, the most rapid change in the V-C dimension occurs when m is small. In the case of $\min VC(m)$, the rapid changes occur both when m is small and when m is large, with little change for $m \approx \frac{1}{2}|\mathcal{F}|$. Notice that for $m = \frac{1}{2}|\mathcal{F}|$, the V-C dimension is bounded between $\lfloor \frac{1}{2} n \rfloor$ and $n - 1$.

3.4 Approximations to the Bounds on the V-C Dimension

The formula for the minimum V-C dimension (equation 3.1) is not in closed form. Unfortunately, a closed form does not exist[GKP89, page 165]. We can, however, approximate the function, as in the following theorem.

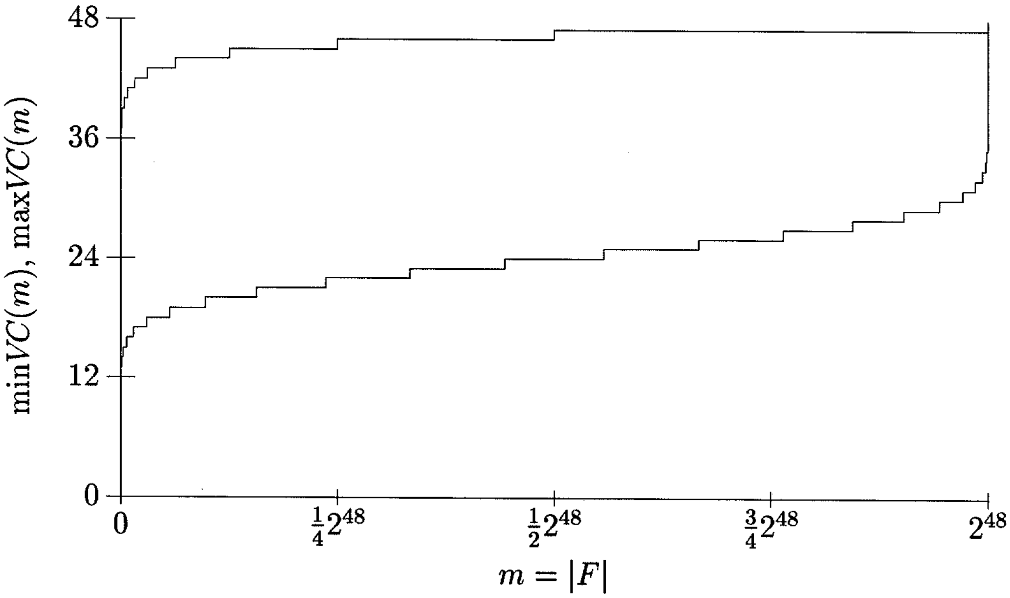

Figure 3.1: min$VC(m)$ and max$VC(m)$, linear scale.

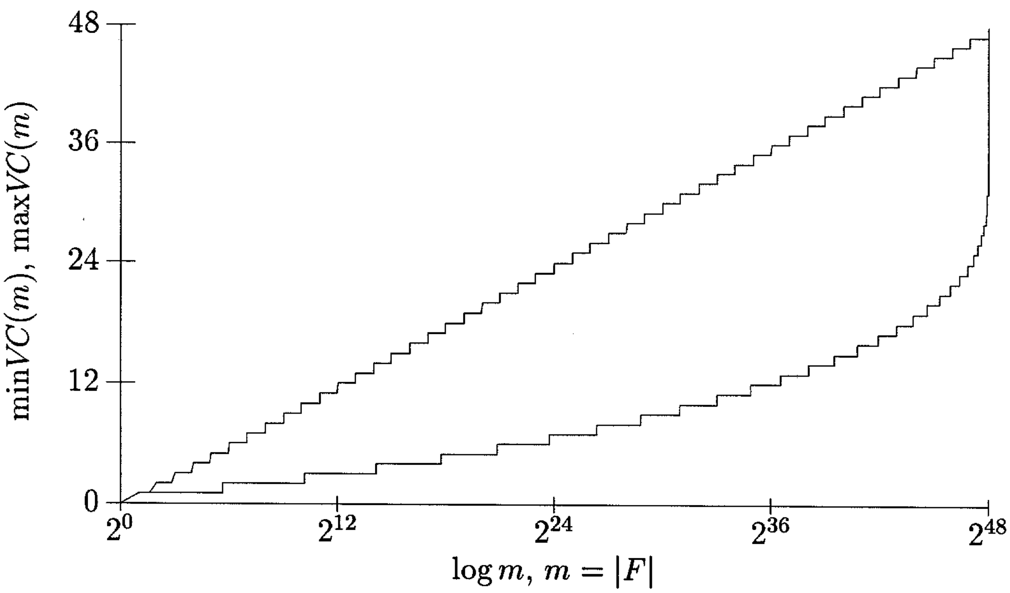

Figure 3.2: min$VC(m)$ and max$VC(m)$, log scale.

Theorem 3.4 For $|F| = m$, $1 \le m < \frac{1}{2}2^n$, $n \ge 3$,

$$\min VC(m) \;\le\; \left\lceil n\mathcal{H}^{-1}\left(\frac{\log_2 m + \frac{1}{2}\log_2 n + \frac{1}{2}}{n}\right)\right\rceil, \qquad (3.3)$$

and

$$\min VC(m) \;>\; \left\lfloor n\mathcal{H}^{-1}\left(\frac{\log_2 m - \log_2 n + 1}{n}\right)\right\rfloor. \qquad (3.4)$$

Here $\mathcal{H}(p) = -p\log_2 p - (1-p)\log_2(1-p)$, the binary entropy function, and we are taking

$$\mathcal{H}^{-1}(x) = \begin{cases} 0 & \text{if } x < 0, \\ y & \text{if } 0 \le x \le 1, \text{ where } x = \mathcal{H}(y) \text{ and } 0 \le y \le \frac{1}{2}, \\ \frac{1}{2} & \text{if } x > 1. \end{cases}$$

Proof: From [PW72, Appendix A] we get the following inequalities:

$$\binom{n}{\lambda n} < \sum_{i=0}^{\lambda n}\binom{n}{i} < \frac{1-\lambda}{1-2\lambda}\binom{n}{\lambda n}, \text{ for } \lambda < \frac{1}{2},$$

and

$$\frac{1}{2}\sqrt{\pi}G \le \binom{n}{\lambda n} < G,$$

where

$$G = \frac{1}{\sqrt{2\pi n\lambda\mu}}\lambda^{-\lambda n}\mu^{-\mu n} = \frac{1}{\sqrt{2\pi n\lambda\mu}}2^{\mathcal{H}(\lambda)n}$$

and $\mu = 1 - \lambda$.

We can combine the two inequalities and get

$$\frac{1}{2}\sqrt{\pi}\frac{1}{\sqrt{2\pi n\lambda\mu}}2^{\mathcal{H}(\lambda)n} < \sum_{i=0}^{\lambda n}\binom{n}{i} < \frac{1-\lambda}{1-2\lambda}\frac{1}{\sqrt{2\pi n\lambda\mu}}2^{\mathcal{H}(\lambda)n}. \qquad (3.5)$$

We will use the above inequality to approximate the function

$$\min VC(m) = \min\left\{ r \;\middle|\; \sum_{i=0}^{r}\binom{n}{i} \ge m \right\}.$$

Suppose we let $r = \lambda n$, and let

$$m = \frac{1}{2}\sqrt{\pi}\frac{1}{\sqrt{2\pi n\lambda\mu}}2^{\mathcal{H}(\lambda)n}$$

(the left-hand side of equation 3.5). Solving for λ in term of m gives us $m < \sum_{i=0}^{\lceil\lambda n\rceil}\binom{n}{i}$; hence, $\min VC(m) \leq \lceil n\lambda\rceil$. Taking logarithms we get

$$\log_2 m = -\frac{1}{2}\log_2(8n\lambda\mu) + \mathcal{H}(\lambda)n.$$

An upper bound for $\lambda\mu = \lambda(1-\lambda)$ is $\frac{1}{4}$, which is attained by $\lambda = \frac{1}{2}$. Thus,

$$\begin{aligned} \log_2 m &> -\frac{1}{2}\log_2(8\cdot n\cdot\frac{1}{4}) + \mathcal{H}(\lambda)n \\ \log_2 m + \frac{1}{2}\log_2 2n &> \mathcal{H}(\lambda)n \end{aligned}$$

and we get

$$\lambda < \mathcal{H}^{-1}\left(\frac{\log_2 m + \frac{1}{2}\log_2 n + \frac{1}{2}}{n}\right)$$

and

$$\min VC(m) \leq \left\lceil n\mathcal{H}^{-1}\left(\frac{\log_2 m + \frac{1}{2}\log_2 n + \frac{1}{2}}{n}\right)\right\rceil.$$

Now, let

$$m = \frac{1-\lambda}{1-2\lambda}\frac{1}{\sqrt{2\pi n\lambda\mu}}2^{\mathcal{H}(\lambda)n}$$

(the right-hand side of equation 3.5). Again solving for λ in terms of m we get $m > \sum_{i=0}^{\lfloor\lambda n\rfloor}\binom{n}{i}$ and $\min VC(m) > \lfloor n\lambda\rfloor$. Taking logarithms, we get

$$\log_2 m = \log_2(1-\lambda) - \log_2(1-2\lambda) - \frac{1}{2}\log_2(2\pi n\lambda\mu) + \mathcal{H}(\lambda)n.$$

An upper bound for $\log_2(1-\lambda)$ is 0. For $\log_2(1-2\lambda)$, first observe that $\lambda n < \frac{1}{2}n$, as $\lambda < \frac{1}{2}$; hence, $\lambda n \leq \frac{1}{2}n - \frac{1}{2}$. From this we get $1 - 2\lambda \geq \frac{1}{n}$, and a lower bound for $\log_2(1-2\lambda)$ is $-\log_2 n$. Finally, notice that $\lambda n \geq 1$; thus $\lambda \geq \frac{1}{n}$. Thus, we get the lower bound $\log_2\left(2\pi\left(1-\frac{1}{n}\right)\right)$ for $\log_2(2\pi n\lambda\mu)$.

Furthermore, if $n \geq 3$, we have $2\pi\left(1-\frac{1}{n}\right) > 4$; hence, we can use 2 as a lower bound for $\log_2(2\pi n\lambda\mu)$. Substituting in the equation above, we get

$$\begin{aligned}\log_2 m &< \log_2 n - \frac{1}{2}\cdot 2 + \mathcal{H}(\lambda)n \\ \log_2 m - \log_2 n + 1 &< \mathcal{H}(\lambda)n.\end{aligned}$$

Then

$$\lambda > \mathcal{H}^{-1}\left(\frac{\log_2 m - \log_2 n + 1}{n}\right),$$

and

$$\min VC(m) > \left\lfloor n\mathcal{H}^{-1}\left(\frac{\log_2 m - \log_2 n + 1}{n}\right)\right\rfloor. \quad \blacksquare$$

In Figure 3.3 we plot the upper and lower bounds on $\min VC(m)$, along with $\min VC(m)$, again for $n = 48$. Since $\min VC(\frac{1}{2}2^n) = \lfloor\frac{n}{2}\rfloor$, the range of the functions is $[0, 24]$.

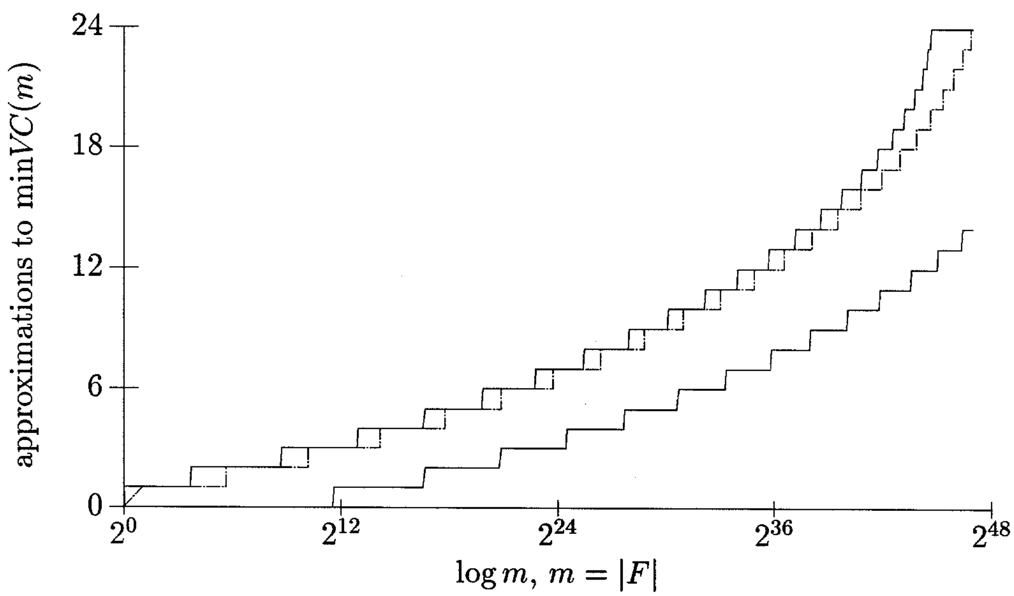

Figure 3.3: Approximations to $\min VC(m)$, log scale.

If we ignore the $\log_2 n$ and constant terms in equations 3.3 and 3.4 we get, for $1 \leq m < \frac{1}{2}2^n$,

$$\begin{aligned} \min VC(m) &\approx n \times \mathcal{H}^{-1}\left(\frac{\log_2 m}{n}\right) \\ \max VC(m) &\approx n \times \left(\frac{\log_2 m}{n}\right). \end{aligned}$$

Chapter 4

Characterizing Sets with $VC(F) = r$

In Theorem 3.1, we constructed a set of functions, F_1, with $|F_1| = \sum_{i=0}^{r} \binom{n}{i}$ and $VC(F_1) = r$. This is the largest set we can have with V-C dimension r, for if $F \subseteq \mathcal{F}$ and $F > \sum_{i=0}^{r} \binom{n}{i}$, then $VC(F) > r$.

In Lemma 3.2, we transformed a set of volumes to another set of volumes that had a minimally sized union. The complement of the union of the volumes defines a set of functions F. Here we have, for $1 \leq r \leq n$, $VC(F) < r$ by construction. The union itself turns out to be the set of all strings with r or more ones, and so the set F is the set of functions whose characteristic strings have less than r ones. Thus, the set of functions, F, is equivalent to the set F_1 constructed in Theorem 3.1, though in this case we have $|F_1| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F_1) = r - 1$.

This set F_1 is not unique in having size $|F_1| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F_1) = r - 1$, $1 \leq r \leq n$. Of course we always get one other, by switching all the ones for zeros and vice versa. There are, however, at least 2^n such sets of functions.

In the proof of Lemma 3.2, we looked at each bit i in turn and set all assignments of that bit to 1. Suppose we have some fixed binary string, $b \in \{0,1\}^n$, and we instead choose to set all assignments of bit i to b_i. From this we obtain a set of functions, $F_{\mathbf{b}}$. If we let $\mathbf{1} = 11\ldots1$, then $F_{\mathbf{1}}$ is equivalent to F_1.

Consider what would happen should we transform, as in Lemma 3.2, $F_{\mathbf{b}}$

into F_1. As was observed after Lemma 3.2, the size of the union does not change if all the assigned bits in a given bit position i are already the same, so we must have $|F_{\boldsymbol{b}}| = |F_1|$. Alternatively, Lemma 3.2 does not require that we change the assignments of bit i to 1. We could have used, in any bit position, 0 instead of 1. Thus, we can use Lemma 3.2 to transform $F_{\boldsymbol{b}}$ into F_1 and back to $F_{\boldsymbol{b}}$. Throughout the transformation the size of the set of functions can only increase or stay the same. Since we end up where we started, it must have been that the set of functions remained the same size throughout the transformation, and again $|F_{\boldsymbol{b}}| = |F_1|$.

Since $|F_{\boldsymbol{b}}| = |F_1|$, we have that $|F_{\boldsymbol{b}}| = \sum_{i=0}^{r-1} \binom{n}{i} > \sum_{i=0}^{r-2} \binom{n}{i}$, hence, $VC(F_{\boldsymbol{b}}) > r - 2$ because of Theorem 3.1. Also $VC(F_{\boldsymbol{b}}) < r$ since we are, by construction, making sure that no r–subset of $\mathcal{X}$ can shatter $F_{\boldsymbol{b}}$. Hence, $VC(F_{\boldsymbol{b}}) = r - 1$.

What we have is a family of sets $F_{\boldsymbol{b}}$, all of which have $|F_{\boldsymbol{b}}| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F_{\boldsymbol{b}}) = r - 1$. Next we will show that if $\boldsymbol{b} \neq \boldsymbol{c}$, then $F_{\boldsymbol{b}} \neq F_{\boldsymbol{c}}$. Given that, we will have at least 2^n largest sets with V-C dimension $r - 1$.

Suppose $r = n$. Then the set ${F_{\boldsymbol{b}}}^c$ contains a single element, say f, and $\mathcal{X}_f = \boldsymbol{b}$. Thus, if $\boldsymbol{b} \neq \boldsymbol{c}$, then $F_{\boldsymbol{b}} \neq F_{\boldsymbol{c}}$.

Now suppose that $1 \leq r < n$. Let $C_{\boldsymbol{b}} = \{ \mathcal{X}_f \mid f \notin F \}$. We have $F_{\boldsymbol{b}} = F_{\boldsymbol{c}}$ if and only if $C_{\boldsymbol{b}} = C_{\boldsymbol{c}}$. Note that the set $C_{\boldsymbol{b}}$ is the same as the set C used in Lemma 3.2 and that $b \in C_{\boldsymbol{b}}$. Suppose that $\boldsymbol{b} \neq \boldsymbol{c}$. If $\boldsymbol{c} \notin C_{\boldsymbol{b}}$, then $F_{\boldsymbol{b}} \neq F_{\boldsymbol{c}}$, as $\boldsymbol{c} \in C_{\boldsymbol{c}}$, so suppose $\boldsymbol{c} \in C_{\boldsymbol{b}}$. The set $C_{\boldsymbol{b}}$ is the union of $\binom{n}{r}$ volumes, and each element of a given volume has a particular r bits in common with $\boldsymbol{b}$. The string $\boldsymbol{c}$ must be in one of these volumes, and therefore must be the same as $\boldsymbol{b}$ in r bit positions. Without loss of generality, assume that the first r of $\boldsymbol{b}$ and $\boldsymbol{c}$ bits are the same. Since $\boldsymbol{b} \neq \boldsymbol{c}$, there must be at least one bit that is different, so assume, again without loss of generality, that it is bit $r + 1$. Thus, we get

$$\begin{aligned} \boldsymbol{b} &= b_0 b_1 \ldots b_{r-1} b_r b_{r+1} b_{r+2} \ldots b_n, \\ \boldsymbol{c} &= b_0 b_1 \ldots b_{r-1} b_r \bar{b}_{r+1} c_{r+2} \ldots c_n. \end{aligned}$$

Consider the string

$$\boldsymbol{b}' = b_0 b_1 \ldots b_{r-1} \bar{b}_r \bar{b}_{r+1} \bar{b}_{r+2} \ldots \bar{b}_n.$$

The string $\boldsymbol{b}'$ is the same as $\boldsymbol{b}$ in only $r - 1$ bits, so $\boldsymbol{b}' \notin C_{\boldsymbol{b}}$. However, the string $\boldsymbol{b}'$ is the same as $\boldsymbol{c}$ in at least r bits, so $\boldsymbol{b}' \in C_{\boldsymbol{c}}$, and $F_{\boldsymbol{b}} \neq F_{\boldsymbol{c}}$.

Thus, we have $F_{\boldsymbol{b}} \neq F_{\boldsymbol{c}}$ if $\boldsymbol{b} \neq \boldsymbol{c}$, hence, at least 2^n distinct sets, $F_{\boldsymbol{b}}$, with $VC(F_{\boldsymbol{b}}) = r - 1$ and $|F_{\boldsymbol{b}}| = \sum_{i=0}^{r-1} \binom{n}{i}$, $1 \leq r \leq n$.

The sets, $F_{\boldsymbol{b}}$, arise another way. Define $\oplus_{\boldsymbol{b}} : \mathcal{F} \to \mathcal{F}$, with $g = \oplus_{\boldsymbol{b}}(f)$ if $\mathcal{X}_g = \mathcal{X}_f \oplus \boldsymbol{b}$, where $\mathcal{X}_f$ is the characteristic string of f, and $\oplus$ is the standard exclusive or operator. Note that $\oplus_{\boldsymbol{b}}(\oplus_{\boldsymbol{b}}(f)) = f$,[1] and that the mapping is a bijection. Let $\oplus_{\boldsymbol{b}}(F) = \{ \oplus_{\boldsymbol{b}}(f) \mid f \in F \}$, and let $\overline{\boldsymbol{b}} = \bar{b}_1\bar{b}_2 \ldots \bar{b}_n$, where $\bar{b}_i = 1 - b_i$, the standard binary complement. Then $\oplus_{\overline{\boldsymbol{b}}}(F_1) = F_{\boldsymbol{b}}$.

First we see that $|\oplus_{\overline{\boldsymbol{b}}}(F_1)| = |F_1| = |F_{\boldsymbol{b}}|$, as the mapping is one to one. To see that $\oplus_{\overline{\boldsymbol{b}}}(F_1) = F_{\boldsymbol{b}}$, consider the volumes, B_i, that make up the complement ${F_1}^c = \mathcal{F} \setminus F_1$ and their transformation, $\oplus_{\overline{\boldsymbol{b}}}(B_i)$. Pick a volume, B_i. We obtain it by taking all the strings that assigned the value one to $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$. The volume $\oplus_{\overline{\boldsymbol{b}}}(B_i)$ is then all strings that assign the value $x_{i_k} \oplus \bar{b}_{i_k} = 1 \oplus \bar{b}_{i_k} = b_{i_k}$ to those x_{i_k}. Thus, the volumes $\oplus_{\overline{\boldsymbol{b}}}(B_i)$ are precisely the volumes that make up ${F_{\boldsymbol{b}}}^c$, and we have $\oplus_{\overline{\boldsymbol{b}}}(F_1) = F_{\boldsymbol{b}}$.

Each volume that makes up ${F_{\boldsymbol{b}}}^c$ contains, by construction, strings that match $\boldsymbol{b}$ in at least r bit positions, hence are Hamming distance at least r from $\overline{\boldsymbol{b}}$. Any string of Hamming distance at least r from $\overline{\boldsymbol{b}}$ must be in at least one volume, namely any volume that corresponds to a selection of r of the bits where the string and $\overline{\boldsymbol{b}}$ differ. Thus, the set of functions $F_{\boldsymbol{b}}$, $VC(F_{\boldsymbol{b}}) = r - 1$, is the set of functions whose characteristic strings are at Hamming distance at most $r - 1$ from the string $\overline{\boldsymbol{b}}$.

These sets $F_{\boldsymbol{b}}$ are not the only subsets $F \subseteq \mathcal{F}$ with $|F| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F) = r - 1$, $1 < r < n$. First, we need the following lemma:

Lemma 4.1 Define $\binom{n}{r}$ volumes $B_i \subseteq \mathcal{B} = \{0,1\}^n$, one for each of the $\binom{n}{r}$ choices of r bits, $1 < r \leq n$. Suppose volume B_i corresponds to the choice of bits $i_1, i_2, \ldots, i_r$, with $i_1 < i_2 < \ldots < i_r$. Then, for volume B_i, we assign 1 to bits $i_1, i_2, \ldots, i_p$, and 0 to bits $i_{p+1}, \ldots, i_r$, where $1 \leq p < r$.

Then, if

$$C = \bigcup_{i=1}^{\binom{n}{r}} B_i,$$

$|C| = \sum_{i=r}^{n} \binom{n}{i}$.

[1]This condition implies that $\oplus_{\boldsymbol{b}}$ is both surjective and injective. First, if $f \in \mathcal{F}$, then $\exists g \in \mathcal{F}$ with $\oplus_{\boldsymbol{b}}(g) = f$, namely $g = \oplus_{\boldsymbol{b}}(f)$. Also, if $\oplus_{\boldsymbol{b}}(f) = \oplus_{\boldsymbol{b}}(g)$, then $\oplus_{\boldsymbol{b}}(\oplus_{\boldsymbol{b}}(f)) = \oplus_{\boldsymbol{b}}(\oplus_{\boldsymbol{b}}(g))$, that is, $f = g$. Thus $\oplus_{\boldsymbol{b}}$ is a bijection.

Proof: Each volume represents strings that have p ones preceding $q = r - p$ zeros. Furthermore, each string of that form appears in some volume. Hence, C is the set off all strings of that form.

First suppose that the p^{th} 1 appears in bit position $(s+1)$. Then we must have exactly $(p-1)$ 1's in the first s bit positions, and at least q zeros in the remaining $(n-s-1)$ bit positions. The following illustrates such a string:

$$\underbrace{\overbrace{\begin{matrix} p-1 & \text{ones} \\ * & \text{zeros} \end{matrix}}^{s \text{ bits}} \; 1 \; \overbrace{\begin{matrix} * & \text{ones} \\ \geq q & \text{zeros} \end{matrix}}^{n-s-1 \text{ bits}}}_{n \text{ bits}}$$

For each choice of s we can sum up the number of such strings by summing over the possible choices for the number of zeros that follow the p^{th} one. Then summing over s gives us the size of the cover as follows:

$$\begin{aligned}
|C| &= \sum_{s=p-1}^{n-1} \binom{s}{p-1} \sum_{k=q}^{n-s-1} \binom{n-s-1}{k} \\
&= \sum_{s=p-1}^{n-1} \binom{s}{p-1} \sum_{k=q}^{n} \binom{n-s-1}{k} \\
&= \sum_{s=p-1}^{n-1} \sum_{k=q}^{n} \binom{s}{p-1} \binom{n-s-1}{k} \\
&= \sum_{k=q}^{n} \sum_{s=p-1}^{n-1} \binom{n-s-1}{k} \binom{s}{p-1} \\
&= \sum_{k=q}^{n} \sum_{t=0}^{n-p} \binom{n-p-t}{k} \binom{p-1+t}{p-1} \\
&= \sum_{k=q}^{n} \binom{n-p+p-1+1}{k+p-1+1} \\
&= \sum_{k=q}^{n} \binom{n}{k+p} \\
&= \sum_{i=p+q}^{n+p} \binom{n}{i} \\
&= \sum_{i=r}^{n} \binom{n}{i}.
\end{aligned}$$

In the above calculation, the double summation was reduced to a single summation using the following identity, which appears in [GKP89, Table 169], and is proven in Appendix A.

$$\sum_{k=0}^{l} \binom{l-k}{m}\binom{q+k}{n} = \binom{l+q+1}{m+n+1}$$

The other steps are straightforward. ∎

For each choice of p in Lemma 4.1, we get a set of functions F with $|F| = \sum_{i=0}^{r-1}\binom{n}{i}$. We have $VC(F) > r-2$ from Theorem 3.1, and $VC(F) < r$, by construction. Hence, $VC(F) = r-1$. It will not necessarily be the case that there is a binary string $\boldsymbol{b}$ with $F = F_{\boldsymbol{b}}$.

First suppose $r = n$, in which case $|F| = \sum_{i=0}^{n-1}\binom{n}{i}$, $VC(F) = n-1$. Then the set F^c contains a single element, say f, and, if $\boldsymbol{b} = \chi_f$, then $F = F_{\boldsymbol{b}}$.

Now suppose that $1 < r < n$, and that $\exists \boldsymbol{b} = b_1 b_2 \ldots b_n$, with $F = F_{\boldsymbol{b}}$. Let $X_1 = \{x_1, x_2, \ldots, x_r\}$ and $X_2 = \{x_2, x_3, \ldots, x_{r+1}\}$. Then, volume B_1 assigns 1 to $x_1, x_2, \ldots, x_p$ and 0 to $x_{p+1}, \ldots, x_r$, while volume B_2 assigns 1 to $x_2, x_3, \ldots, x_{p+1}$ and 0 to $x_{p+2}, \ldots, x_{r+1}$. All binary strings in both B_1 and B_2 must be Hamming distance at least r from $\overline{\boldsymbol{b}}$. Since

$$\boldsymbol{b}_1 = \underbrace{1\ldots1}_{p \text{ bits}}\underbrace{0\ldots0}_{r-p \text{ bits}}\bar{b}_{r+1}\bar{b}_{r+2}\ldots\bar{b}_n \in B_1,$$

we must have

$$\overline{\boldsymbol{b}} = \underbrace{0\ldots0}_{p \text{ bits}}\underbrace{1\ldots1}_{r-p \text{ bits}}\bar{b}_{r+1}\bar{b}_{r+2}\ldots\bar{b}_n$$

in order to ensure that the Hamming distance between $\boldsymbol{b}_1$ and $\overline{\boldsymbol{b}}$ is at least r. Now, let

$$\boldsymbol{b}_2 = 0\underbrace{1\ldots1}_{p \text{ bits}}\underbrace{0\ldots0}_{r-p \text{ bits}}\bar{b}_{r+2}\ldots\bar{b}_n \in B_2.$$

The strings $\overline{\boldsymbol{b}}$ and $\boldsymbol{b}_2$ agree in bit positions $1, p+1, r+2, r+3, \ldots n$, or in $2+(n-(r+2)+1) = n-r+1$ bit positions. Hence, the Hamming distance between $\boldsymbol{b}_2$ and $\overline{\boldsymbol{b}}$ is at most $r-1$, contradicting the requirement that all elements of B_2, including $\boldsymbol{b}_2$, be at least Hamming distance r from $\overline{\boldsymbol{b}}$. Thus, $F \neq F_{\boldsymbol{b}}$ for any string $\boldsymbol{b}$.

Thus, we have shown that the set of functions $F_{\boldsymbol{b}}$ with Hamming distance at most r from the string $\overline{\boldsymbol{b}}$ do not fully categorize the sets of functions, F, with $|F| = \sum_{i=0}^{r-1}\binom{n}{i}$ and $VC(F) = r-1$.

Chapter 5

The V-C Dimension of a Set and Its Complement

In Theorem 3.4 we developed an approximation to $\min VC(m)$, $m = |F|$, provided that $m < \frac{1}{2}2^n$. If $m = \frac{1}{2}2^n$, we can use equation 3.1 directly to find $\min VC(\frac{1}{2}2^n) = \lfloor \frac{1}{2}n \rfloor$. If $m > \frac{1}{2}2^n$, then $|F^c| = 2^n - m < \frac{1}{2}2^n$. The following corollary relates the minimum V-C dimension of F and the minimum V-C dimension of F^c. This, together with equation 3.1, allows us to approximate the minimum V-C dimension for all m.

Corollary 5.1 Suppose $1 \le m < 2^n$. If $m = \sum_{i=0}^{r} \binom{n}{i}$ for some $r < n$, then

$$\min VC(m) + \min VC(2^n - m) = n - 1. \tag{5.1}$$

Otherwise,

$$\min VC(m) + \min VC(2^n - m) = n. \tag{5.2}$$

These two equations can be reduced to

$$\min VC(m) + \min VC(2^n - m + 1) = n \tag{5.3}$$

for $1 \le m \le 2^n$.

Proof: Suppose $m = \sum_{i=0}^{r} \binom{n}{i}$, for some $r < n$. Then,

$$2^n - m = \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r} \binom{n}{i}$$

$$
\begin{aligned}
&= \sum_{i=r+1}^{n} \binom{n}{i} \\
&= \sum_{j=0}^{n-(r+1)} \binom{n}{n-j} \\
&= \sum_{j=0}^{n-(r+1)} \binom{n}{j}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\min VC(m) + \min VC(2^n - m) &= r + (n - r - 1) \\
&= n - 1.
\end{aligned}
$$

Otherwise, suppose $m \neq \sum_{i=0}^{r} \binom{n}{i}$, $1 < m < 2^n$. Then there is an r such that $\sum_{i=0}^{r-1} \binom{n}{i} < m$ and $\sum_{i=0}^{r} \binom{n}{i} > m$, and

$$
\begin{aligned}
2^n - m &> \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r} \binom{n}{i} \\
&= \sum_{i=r+1}^{n} \binom{n}{i} \\
&= \sum_{i=0}^{n-(r+1)} \binom{n}{i},
\end{aligned}
$$

and

$$
\begin{aligned}
2^n - m &< \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r-1} \binom{n}{i} \\
&= \sum_{i=r}^{n} \binom{n}{i} \\
&= \sum_{i=0}^{n-r} \binom{n}{i}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\min VC(m) + \min VC(2^n - m) &= r + (n - r) \\
&= n.
\end{aligned}
$$

This verifies equations 5.1 and 5.2. For equation 5.3, first observe that if $m = 1$, then

$$\begin{aligned} \min VC(m) + \min VC(2^n - m + 1) &= 0 + n \\ &= n. \end{aligned}$$

Suppose that $1 < m \leq 2^n$, and let r be such that $\sum_{i=0}^{r-1} \binom{n}{i} < m$ and $\sum_{i=0}^{r} \binom{n}{i} \geq m$. Then

$$\begin{aligned} 2^n - m + 1 &\geq \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r} \binom{n}{i} + 1 \\ &> \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r} \binom{n}{i} \\ &= \sum_{i=r+1}^{n} \binom{n}{i} \\ &= \sum_{i=0}^{n-(r+1)} \binom{n}{i}, \end{aligned}$$

and

$$\begin{aligned} 2^n - m + 1 &< \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r-1} \binom{n}{i} + 1 \\ &\leq \sum_{i=0}^{n} \binom{n}{i} - \sum_{i=0}^{r-1} \binom{n}{i} \\ &= \sum_{i=r}^{n} \binom{n}{i} \\ &= \sum_{i=0}^{n-r} \binom{n}{i}. \end{aligned}$$

Here we get

$$\begin{aligned} \min VC(m) + \min VC(2^n - m + 1) &= r + (n - r) \\ &= n. \blacksquare \end{aligned}$$

Consider once again the set F_1. We have $|F_1| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F_1) = r-1$. The set F_1 is the set of all functions whose characteristic strings have at most $(r-1)$ 1's. Thus the set ${F_1}^c$ is the set of all functions whose characteristic strings have at least r 1's, or, equivalently, at most $(n-r)$ 0's. Thus, we have $|{F_1}^c| = \sum_{i=0}^{n-r} \binom{n}{i}$, and $VC({F_1}^c) = n-r$. Both $VC(F_1) = \min VC(|F_1|)$ and $VC({F_1}^c) = \min VC(|{F_1}^c|)$; therefore we have an example where the lower bound represented by equation 5.1 is attained.

Suppose instead of dividing $\mathcal{F}$ into two sets, F_1 and ${F_1}^c$, we divide $\mathcal{F}$ into three disjoint sets as follows:

$$
\begin{aligned}
F &= \{\, f \mid \chi_f \text{ has} < r \text{ ones} \,\} \\
G &= \{\, f \mid \chi_f \text{ has exactly } r \text{ ones} \,\} \\
H &= \{\, f \mid \chi_f \text{ has} > r \text{ ones} \,\}
\end{aligned}
$$

Notice that, as above, $|F| = \sum_{i=0}^{r-1} \binom{n}{i}$ and $VC(F) = r-1$, and $|F \cup G| = \sum_{i=0}^{r} \binom{n}{i}$ and $VC(F \cup G) = r$. Similarly, $|H| = \sum_{i=0}^{n-r-1} \binom{n}{i}$ and $VC(H) = n-r-1$, and $|H \cup G| = \sum_{i=0}^{n-r} \binom{n}{i}$ and $VC(H \cup G) = n-r$. Now, suppose that we divide G into two non-empty, disjoint sets, G_1 and G_2. Then we have $\sum_{i=0}^{r-1} \binom{n}{i} < |F \cup G_1| < \sum_{i=0}^{r} \binom{n}{i}$, and $VC(F \cup G_1) = r$. The set $H \cup G_2$ is the complement of $F \cup G_1$, and we have $\sum_{i=0}^{n-r-1} \binom{n}{i} < |H \cup G_2| < \sum_{i=0}^{n-r} \binom{n}{i}$, and $VC(H \cup G_2) = n-r$. Hence we have shown that the lower bound represented by equation 5.2 is also attained.

In a completely analogous way to the above, we can demonstrate that the lower bound represented by equation 5.3 is likewise attained.

The example attaining the lower bound for equation 5.1 is noteworthy because it provides an example of two sets F_1 and ${F_1}^c$ with

$$VC(F_1 \cup {F_1}^c) > VC(F_1) + VC({F_1}^c).$$

This is very different from what we would expect if the V-C dimension were equal to $\max VC(m)$, as the upper bound is logarithmic in the size of the sets.

Chapter 6

Kolmogorov Complexity and the V-C Dimension

In this chapter we will be using Turing machines to define Kolmogorov complexity[Cha87, KU]. We will deal informally here with Turing machines. A more rigorous treatment of these machines can be found in such references such as [HU79].

A Turing machine has two parts: a finite state control, and a read/write infinite tape. A Turing machine starts with an input that is written on the tape and the control directed to it. If the machine eventually halts with a single string left on the tape, then that string is considered to be the output of the machine.

The class of Turing machines is *universal* in that there exists a Turning machine U that, when given a description of any Turning machine, M, and an input string, $\boldsymbol{b}$, gives the same output as M would if given the input $\boldsymbol{b}$. If we let $\rho(M)$ be a self-delimiting string that encodes the machine M, then we can provide the string $\rho(M)\boldsymbol{b}$ as input to U. This is equivalent, in terms of input/output behaviour, to having given $\boldsymbol{b}$ as input to M.

The Kolmogorov complexity of a string, $\boldsymbol{c}$, is the length of the shortest input to U that produces $\boldsymbol{c}$ as output, or

$$K(\boldsymbol{c}) = \min\{\, |\rho(M)\boldsymbol{b}| \mid U \text{ given } \rho(M)\boldsymbol{b} \text{ outputs } \boldsymbol{c}\,\}.$$

Suppose E is the *everhalting* machine, that is, the machine that immediately halts, leaving the input string $\boldsymbol{b}$ as the output. The machine, E, simply computes the identity function. Then the string $\rho(E)\boldsymbol{c}$, when given to U, outputs $\boldsymbol{c}$; hence, $K(\boldsymbol{c}) \leq |\rho(E)\boldsymbol{c}| = |\rho(E)| + |\boldsymbol{c}| = c + |\boldsymbol{c}|$, where $c = |\rho(E)|$ is a constant. Thus,

$$\forall \boldsymbol{c},\ K(\boldsymbol{c}) \leq |\boldsymbol{c}| + c.$$

We will define the Kolmogorov complexity of a boolean function to be the Kolmogorov complexity of its characteristic string. Consider the set

$$F = \{\, f \mid K(f) \leq r \,\}.$$

Then,

$$VC(F) \geq r - c'$$

for some constant c'.

Consider a string, $\boldsymbol{b}$, of length p. We know that $K(\boldsymbol{b}) \leq p + c$. If we assume that the number n is available to our Turning machine, and therefore need not be specified as part of the input, then we can alter the machine to pad out the string, $\boldsymbol{b}$, to be a string, $\mathbf{c}$, of length n by adding $n - r$ zeros to the end. Such an alteration will add at most a constant to the length of the machine. Thus, we have $K(\boldsymbol{c}) \leq p + c'$. All of these strings will be characteristic strings for functions f in the set F, provided that $p + c' \leq r$. Then, if $p \leq r - c'$, there is a p–subset of $\mathcal{X}$ that shatters F. Thus, $VC(F) \geq p$ for all $p \leq r - c'$; hence, $VC(F) \geq r - c'$.

Chapter 7

Conclusions

We have developed an upper and lower bound on the V-C dimension of a set of functions, F, given only its size, m. Both bounds are attainable, as for any m there is a set F with $|F| = m$ and $VC(F) = \min VC(F)$, and a set G with $|G| = m$ and $VC(G) = \max VC(G)$.

With these bounds we have begun to investigate the properties of the V-C dimension. Future directions include a better characterization of sets with a given V-C dimension, a better understanding of how the V-C dimension behaves under operations such as union and set subtraction, and an investigation of the relationships between the V-C dimension and other measures of complexity, such as Kolmogorov complexity.

Appendix A

A Binomial Identity

Theorem A.1 For $l, m \geq 0$, $n \geq q \geq 0$,

$$\sum_{k=0}^{l} \binom{l-k}{m} \binom{q+k}{n} = \binom{l+q+1}{m+n+1}. \tag{A.1}$$

Proof: To prove this we will use the following identities:

$$\binom{r}{k} = (-1)^k \binom{k-r-1}{k}, \tag{A.2}$$

and

$$\sum_{k=0}^{n} \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}. \tag{A.3}$$

Here we have r, s real numbers, and k, n non-negative integers.

The first identity A.2 follows directly from the definition of $\binom{r}{k}$, namely

$$\binom{r}{k} = \frac{r \cdot (r-1) \cdot \cdots \cdot (r-k+1)}{k!}.$$

We take $\binom{r}{k} = 1$ for all r if $k = 0$.

The second is Vandermonde's convolution. For r and s positive integers, the convolution has a combinatorial interpretation. Suppose we have r red balls and s blue balls. Each term in the left hand summation of A.3 is the number of ways of choosing k red balls and $n-k$ blue balls. Summed over

$0 \leq k \leq n$ we get the number of ways of selecting n balls from among the r red balls and s blue balls, which is precisely the right-hand side of A.3. Now suppose r and s are variables over the real numbers. Then both the left and right sides of A.3 become polynomials in r and s. Both sides are of finite degree, as all terms $a_{ij}r^i s^j$ have $0 \leq i, j \leq n$. Thus, the difference of the two polynomials is also of finite degree. Yet the difference is zero for all positive integers r and s. This implies that the difference polynomial must be zero along any line $r = k$, for every positive integer k, and thus zero along any line $s = x$, for every real x; this makes the polynomial identically zero. Hence, we have Vandermonde's convolution for all real r and s.

In equation A.1, the factor $\binom{l-k}{m}$ is zero if $l - k < m$. Thus we must have $l - k \geq m$, that is, $k \leq l - m \leq l$. Similarly, from the other factor $\binom{q+k}{n}$, we must have $q + k \geq n$, that is, $k \geq n - q \geq 0$, as $n \geq q$. Thus, we can restrict the range of the summation from $0 \leq k \leq n$ to $n - q \leq k \leq l - m$.

$$
\begin{aligned}
&\sum_{k=0}^{l} \binom{l-k}{m}\binom{q+k}{n} \\
&= \sum_{k=n-q}^{l-m} \binom{l-k}{m}\binom{q+k}{n} \\
&= \sum_{k=n-q}^{l-m} \binom{l-k}{l-k-m}\binom{q+k}{q+k-n} \\
&= \sum_{k=n-q}^{l-m} (-1)^{l-k-m}\binom{-m-1}{l-k-m}(-1)^{q+k-n}\binom{-n-1}{q+k-n} \\
&= (-1)^{q+l-m-n} \sum_{k=n-q}^{l-m} \binom{-m-1}{l-k-m}\binom{-n-1}{q+k-n} \\
&= (-1)^{q+l-m-n} \sum_{k=n-q}^{l-m} \binom{-m-1}{l-k-m}\binom{-n-1}{(q+l-m-n)-(l-k-m)} \\
&= (-1)^{q+l-m-n} \sum_{l-k-m=0}^{q+l-m-n} \binom{-m-1}{l-k-m}\binom{-n-1}{(q+l-m-n)-(l-k-m)} \\
&= (-1)^{q+l-m-n} \binom{-m-1-n-1}{q+l-m-n}
\end{aligned}
$$

$$
\begin{aligned}
&= (-1)^{q+l-m-n}(-1)^{q+l-m-n}\binom{(q+l-m-n)-(-m-n-2)-1}{q+l-m-n} \\
&= \binom{q+l+1}{q+l-m-n} \\
&= \binom{q+l+1}{m+n+1}. \blacksquare
\end{aligned}
$$

Bibliography

[AM89] Yaser S. Abu-Mostafa. The Vapnik-Chervonenkis dimension: Information versus complexity in learning. *Neural Computation*, 1:312–317, 1989.

[Cha87] G. J. Chaitin. *Information, Randomness, and Incompleteness.* World Scientific, 1987.

[GKP89] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science.* Addison-Wesley, 1989.

[HU79] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

[KU88] A. N. Kolmogorov and V. A. Uspenskii. Algorithms and randomness. *Theory Probab. Appl.*, 32(3):389–412, 1988.

[PW72] W. Wesley Peterson and E. J. Weldon, Jr. *Error Correcting Codes.* MIT Press, second edition, 1972.

[VC71] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, XVI(2):264–280, 1971.