On the Complexity of Binary Samples

Joel Ratsaby
Electrical and Electronics Engineering Department
Ariel University Center of Samaria
ISRAEL
ratsaby@ariel.ac.il

August 8, 2008

Abstract

Consider a class $\mathcal{H}$ of binary functions $h : X \to \{-1,+1\}$ on an interval $X = [0,B] \subset \mathbb{R}$. Define the *sample width* of h on a finite subset (a sample) $S \subset X$ as $\omega_S(h) = \min_{x\in S} |\omega_h(x)|$ where $\omega_h(x) = h(x)\max\{a \geq 0 : h(z) = h(x), x-a \leq z \leq x+a\}$. Let $\mathbb{S}_\ell$ be the space of all samples in X of cardinality ℓ and consider sets of wide samples, i.e., *hypersets* which are defined as $A_{\beta,h} = \{S \in \mathbb{S}_\ell : \omega_S(h) \geq \beta\}$. Through an application of the Sauer-Shelah result on the density of sets an upper estimate is obtained on the growth function (or trace) of the class $\{A_{\beta,h} : h \in \mathcal{H}\}$, $\beta > 0$, i.e., on the number of possible dichotomies obtained by intersecting all hypersets with a fixed collection of samples $S \in \mathbb{S}_\ell$ of cardinality m. The estimate is $2\sum_{i=0}^{2\lfloor B/(2\beta)\rfloor}\binom{m-\ell}{i}$.

Keywords: Binary functions, density of sets, VC-dimension
AMS Subject Classification: 06E30, 68Q32, 68Q25, 03C13, 68R05

1 Overview

Let $B > 0$ and define the domain as $X = [0,B]$. In this paper we consider the class $\mathcal{H}$ of all *binary functions* $h : X \to \{-1,+1\}$ which have only simple discontinuities, i.e., at any point x the limits $h(x^+) = \lim_{z\to x^+} h(z)$ from the right and similarly from the left $h(x^-)$ exist (but are not necessarily equal). A main theme of our recent work has been to characterize binary functions based on their behavior on a finite subset of X. In Anthony and Ratsaby [2006] it was shown that the problem of learning binary functions from a finite labeled sample can improve the generalization error-bounds if the learner obtains a hypothesis which in addition to minimizing the empirical sample-error is also 'smooth' around elements of the sample. This notion of smoothness is based on the simple notion of *width* of h at x which is defined by

$$\omega_h(x) = h(x)\max\{a \geq 0 : h(z) = h(x), x-a \leq z \leq x+a\}.$$

Viewing a binary function as a decision rule (or classifier), then the width of h on x may represent a form of 'confidence' that h has in the value that it assigns to x since the larger the width at x the farther the region where h decides the opposite.

For a finite subset $S \subset X$ (S is also referred to as a *sample*) the *sample width* of h denoted $\omega_S(h)$ is defined as

$$\omega_S(h) = \min_{x \in S} |\omega_h(x)|.$$

This definition of width resembles the notion of sample *margin* of a real-valued function f (see for instance Anthony and Bartlett [1999]). We say that a sample S is *wide* for h if the width $\omega_S(h)$ is large. Wide samples implicitly contain more side information for instance about a learning problem. The current paper aims at estimating the complexity of the class of wide samples for functions in $\mathcal{H}$. This complexity is related to a notion of description complexity and knowing it enables to compute the efficiency of information that is implicit in samples for learning (see Ratsaby [2007]).

2 Introduction

For any logical expression A denote by $\mathbb{I}\{A\}$ the indicator function which takes the value 1 or 0 whenever the statement A is true or false, respectively. Let ℓ be any fixed positive integer. We denote by $\mathbb{S}_\ell$ the space of all samples $S \subset X$ of size ℓ. On $\mathbb{S}_\ell$, consider the following sets of wide samples,

$$A_{\beta,h} = \{S \in \mathbb{S}_\ell : \omega_S(h) \geq \beta\}, \quad \beta > 0.$$

We refer to such sets as *hypersets*. It will be convenient to associate with these sets the indicator functions on $\mathbb{S}_\ell$ which are denoted by

$$h'_{\beta,h}(S) = \mathbb{I}_{A_{\beta,h}}(S).$$

These are referred to as *hyperconcepts* and we may write h' for brevity. For any fixed width parameter $\gamma > 0$, define the *hyperclass*

$$\mathcal{H}'_\gamma = \left\{h'_{\gamma,h} : h \in \mathcal{H}\right\}. \tag{1}$$

In words, $\mathcal{H}'_\gamma$ consists of all sets of subsets $S \subset X$ of cardinality ℓ on which the corresponding binary functions h are wide by at least γ.

The aim of the paper is to compute the complexity of the hyperclass $\mathcal{H}'_\gamma$ that corresponds to the class $\mathcal{H}$. Since the domain X is infinite, then so is $\mathcal{H}'_\gamma$, and hence, one cannot simply measure its cardinality. Instead, we apply a standard combinatorial measure of the complexity of a family of sets as follows: let Y be an arbitrary domain and $\mathcal{G}$ an infinite class of subsets of Y. For any subset $S = \{y_1, \ldots, y_n\} \subset Y$ let

$$\Gamma_{\mathcal{G}}(S) = |\mathcal{G}_{|S}| \tag{2}$$

where $\mathcal{G}_{|S} = \{[\mathbb{I}_G(y_1), \ldots, \mathbb{I}_G(y_n)] : G \in \mathcal{G}\}$. The vector $[\mathbb{I}_G(y_1), \ldots, \mathbb{I}_G(y_n)]$ is sometimes referred to as a labeling or a dichotomy of the sample S. The *growth function* (see for instance Anthony and Bartlett [1999]) is defined as

$$\Gamma_{\mathcal{G}}(n) = \max_{S: S \subset Y, |S| = n} \Gamma_{\mathcal{G}}(S).$$

It measures the rate at which the number of dichotomies, obtained by intersecting subsets G of $\mathcal{G}$ with a finite set S, increases with respect to the cardinality n of S in the extreme case, i.e., for the maximal possible S. This quantity is also known as the trace of $\mathcal{G}$, see for instance Bollobás [1986].

Since we are interested in hypersets as opposed to simple sets G (as above) then we consider the trace on a finite collection $\zeta \subset \mathbb{S}_\ell$ of samples (instead of a finite sample S as above). It will be convenient to define the cardinality of such a collection as the cardinality of the union of its component sets, i.e., for any given finite collection $\zeta \subset \mathbb{S}_\ell$ let

$$|\zeta| = \left| \bigcup_{S:S\in\zeta} S \right| \tag{3}$$

and we use m to denote a possible value of $|\zeta|$. As a measure of complexity of $\mathcal{H}'_\gamma$ we compute the growth as a function of m, i.e.

$$\Gamma_{\mathcal{H}'_\gamma}(m) = \max_{\zeta:\zeta\subset\mathbb{S}_\ell, |\zeta|=m} \Gamma_{\mathcal{H}'_\gamma}(\zeta).$$

3 Main result

Let us state the main result of the paper.

Theorem 1 *Let $\ell, m > 0$ be integers and $B > 0$ a real number. Let $\mathcal{H}$ be the class of binary functions on $[0, B]$ with only simple discontinuities. For a given parameter value $\gamma > 0$, the growth function of the hyperclass $\mathcal{H}'_\gamma$ on the space $\mathbb{S}_\ell$ satisfies the following upper bound,*

$$\Gamma_{\mathcal{H}'_\gamma}(m) \leq 2 \sum_{i=0}^{2\lfloor B/(2\gamma)\rfloor} \binom{m-\ell}{i}.$$

Remark 1 *For $m > \ell + B/\gamma$, the following simpler bound holds*

$$\Gamma_{\mathcal{H}'_\gamma}(m) \leq 2 \left(\frac{e\gamma(m-\ell)}{B} \right)^{\frac{B}{\gamma}}.$$

Before proving this result we need some additional notation. We denote by $\langle a, b\rangle$ a generalized interval set of the form $[a, b]$, (a, b), $[a, b)$ or $(a, b]$. For a set R we write $\mathbb{I}_R(x)$ to represent the indicator function of the statement $x \in R$. In case of an interval set $R = \langle a, b\rangle$ we write $\mathbb{I}\langle a, b\rangle$.

Proof: Any binary function h may be represented by thresholding a real-valued function f on X, i.e., $h(x) = \text{sgn}(f(x))$ where for any $a \in \mathbb{R}$, $\text{sgn}(a) = +1$ or -1 if $a > 0$ or $a \leq 0$, respectively. The idea is to choose a class $\mathcal{F}$ of real-valued functions f which is rich enough (it has to be infinite since there are infinitely many binary functions on X) but as simple as possible. This is important since, as we will show, the growth function of $\mathcal{H}'_\gamma$ is bounded from above by the complexity of a class that is a variant of $\mathcal{F}$.

We start by constructing such an $\mathcal{F}$. For a binary function h on X consider the corresponding set sequence $\{R_i\}_{i=1,2,\ldots}$ which satisfies the following properties: (a) $[0, B] =$

$\bigcup_{i=1,2,\dots} R_i$ and for any $i \neq j$, $R_i \cap R_j = \emptyset$, (b) h alternates in sign over consecutive sets R_i, R_{i+1}, (c) R_i is an interval set $\langle a, b\rangle$ with possibly $a = b$ (in which case $R_i = \{a\}$). Hence h has the following general form

$$h(x) = \pm \sum_{i=1,2,\dots,} (-1)^i \mathbb{I}_{R_i}(x). \tag{4}$$

Thus there are exactly two functions h corresponding uniquely to each sequence of sets R_i, $i = 1, 2, \dots$.. Unless explicitly specified, the end points of $X = [0, B]$ are not considered roots of h, i.e., the function 'continues' with the same value it takes at the endpoints (formally, $h(x) = h(0)$ for $x < 0$ and $h(x) = h(B)$ for $x > B$). Now, associate with the set sequence $R_1, R_2, \dots$ the unique non-decreasing sequence of right-endpoints $a_1, a_2, \dots$ which define these sets (the sequence may have up to two consecutive repetitions except for 0 and B) according to

$$R_i = \langle a_{i-1}, a_i\rangle,\ i = 1, 2, \dots. \tag{5}$$

with the first left end point being $a_0 = 0$. Note that different choices for $\langle$ and $\rangle$ (see earlier definition of a generalized interval $\langle a, b\rangle$) give different sets R_i and hence different functions h. For instance, suppose that $X = [0, 7]$ then the following set sequence $R_1 = [0, 2.4)$, $R_2 = [2.4, 3.6)$, $R_3 = [3.6, 3.6] = \{3.6\}$, $R_4 = (3.6, 7]$ has a corresponding end-point sequence $a_1 = 2.4, a_2 = 3.6, a_3 = 3.6, a_4 = 7$. Note that a singleton set introduces a repeated value in this sequence. As another example consider $R_1 = [0, 0] = \{0\}$, $R_2 = (0, 4.1)$, $R_3 = [4.1, 7]$ with $a_1 = 0$, $a_2 = 4.1$, $a_3 = 7$.

Next, define the corresponding sequence of midpoints

$$\mu_i = \frac{a_i + a_{i+1}}{2},\ i = 1, 2, \dots.$$

Define the continuous real-valued function $f : X \to [-B, B]$ that corresponds to h (via the end-point sequence) as follows:

$$f(x) = \pm \sum_{i=1,2,\dots} (-1)^{i+1}(x - a_i)\mathbb{I}[\mu_{i-1}, \mu_i] \tag{6}$$

where we take $\mu_0 = 0$ (see for instance, Figure 1). Clearly, the value $f(x)$ equals the width $\omega_h(x)$. Note that for a fixed sequence of endpoints a_i, $i = 1, 2, \dots$ the function f is invariant to the type of intervals $R_i = \langle a_{i-1}, a_i\rangle$ that h has, for instance, the set sequence $[0, a_1)$, $[a_1, a_2)$, $[a_2, a_3]$, $(a_3, B]$ and the sequence $[0, a_1]$, $(a_1, a_2]$, $(a_2, a_3]$, $(a_3, B]$ yield different binary functions h but the same width function f. For convenience, when h has a finite number n of interval sets R_i, then the sum in (4) has an upper limit of n and we define $a_n = B$. Similarly, the sum in (6) goes up to $n - 1$ and we define $\mu_{n-1} = B$. Let us denote by

$$\mathcal{F}_+ = \{|f| : f \in \mathcal{F}\}. \tag{7}$$

It follows that the hyperclass $\mathcal{H}'_\gamma$ may be represented in terms of the class $\mathcal{F}_+$ as follows: define the hypersets

$$A_{\beta,f} = \{S \in \mathbb{S}_\ell : f(x) \geq \beta, x \in S\}, \qquad \beta > 0, f \in \mathcal{F}_+$$

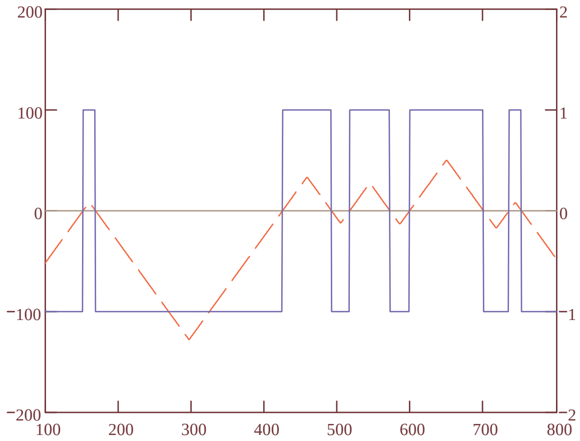

Figure 1: h (solid) and its corresponding f (dashed) on $X = [0, B]$ with $B = 800$

with corresponding hyperconcepts $f'_{\gamma,f} = \mathbb{I}_{A_{\beta,f}}(S)$, let

$$\mathcal{F}'_\gamma = \{f'_{\gamma,f} : f \in \mathcal{F}_+\}$$

and let us define

$$\mathcal{H}'_\gamma = \mathcal{F}'_\gamma. \tag{8}$$

Hence, it suffices to compute the growth function $\mathit{\Gamma}_{\mathcal{F}'_\gamma}(m)$.

Let us now begin to analyze the hyperclass $\mathcal{F}'_\gamma$. By definition, $\mathcal{F}'_\gamma$ is a class of indicator functions of subsets of $\mathbb{S}_\ell$. Denote by $\zeta_N \subset \mathbb{S}_\ell$ a collection of N such subsets. By a *generalized* collection we will mean a collection of subsets $S \subset X$ with cardinality $|S| \leq \ell$. Henceforth we fix a value m and consider only collections

$$\zeta_N, \text{ such that } |\zeta_N| = m \tag{9}$$

where the definition of cardinality is according to (3). Let us denote the individual components of ζ_N by $S^{(j)} \in \mathbb{S}_\ell$, $1 \leq j \leq N$ hence

$$\zeta_N = \{S^{(1)}, \ldots, S^{(N)}\}.$$

The growth function may be expressed as

$$\mathit{\Gamma}_{\mathcal{F}'_\gamma}(m) = \max_{\zeta_N \subset \mathbb{S}_\ell, |\zeta_N| = m} \mathit{\Gamma}_{\mathcal{F}'_\gamma}(\zeta_N) = \max_{\zeta_N \subset \mathbb{S}_\ell, |\zeta_N| = m} \left| \left\{ [f'(S^{(1)}), \ldots, f'(S^{(N)})] : f' \in \mathcal{F}'_\gamma \right\} \right|. \tag{10}$$

Denote by $S_i^{(j)}$ the i^{th} element of the sample $S^{(j)}$ based on the ordering of the elements of $S^{(j)}$ (which is induced by the ordering on X). Then

$$\begin{aligned} &\mathit{\Gamma}_{\mathcal{F}'_\gamma}(\zeta_N) \\ &= \left| \left\{ \left[\mathbb{I}\left(\min_{x \in S^{(1)}} f(x) > \gamma \right), \ldots, \mathbb{I}\left(\min_{x \in S^{(N)}} f(x) > \gamma \right) \right] : f \in \mathcal{F}_+ \right\} \right| \\ &= \left| \left\{ \left[\prod_{j=1}^{\ell} \mathbb{I}\left(f(S_j^{(1)}) > \gamma \right), \ldots, \prod_{j=1}^{\ell} \mathbb{I}\left(f(S_j^{(N)}) > \gamma \right) \right] : f \in \mathcal{F}_+ \right\} \right|. \end{aligned} \tag{11}$$

Order the elements in each component of ζ_N by the underlying ordering on X. Then put the sets in lexical ordering starting with the first up to the ℓ^{th} element. For instance, suppose $m = 7$, $N = 3$, $\ell = 4$ and

$$\zeta_3 = \{ \ \{2,8,9,10\}, \{2,5,8,9\}, \{3,8,10,13\}\}$$

then the ordered version is

$$\{\{2,5,8,9\}, \{2,8,9,10\}, \{3,8,10,13\}\}.$$

For any $x \in X$ let

$$\theta_f^{\gamma}(x) = \mathbb{I}\left(f(x) > \gamma\right) \tag{12}$$

(we will sometimes write $\theta_f(x)$ for short). For any sample $S^{(i)}$ of cardinality $|S^{(i)}| \geq 1$ let

$$e_{S^{(i)}}(f) = \prod_{j=1}^{|S^{(i)}|} \theta_f(S_j^{(i)}).$$

Then for ζ_N we denote by

$$v_{\zeta_N}(f) = [e_{S^{(1)}}(f), \ldots, e_{S^{(N)}}(f)]$$

where for brevity we sometimes write $v(f)$. Let

$$V_{\mathcal{F}_+}(\zeta_N) = \{v_{\zeta_N}(f) : f \in \mathcal{F}_+\}$$

or simply $V(\zeta_N)$. Then from (11) we have

$$\varGamma_{\mathcal{F}'_\gamma}(\zeta_N) = \left|V_{\mathcal{F}_+}(\zeta_N)\right|. \tag{13}$$

Denote by X' the union

$$X' = \bigcup_{j=1}^{N} S^{(j)} = \{x_i\}_{i=1}^m \subset X \tag{14}$$

and take the elements to be ordered as $x_i < x_{i+1}$, $1 \leq i \leq m-1$. The dependence of X' on ζ_N is left implicit. We will need the following procedure which maps ζ_N to a generalized collection.

Procedure G: *Given ζ_N construct $\zeta_{\hat{N}}$ as follows: Let $\hat{S}^{(1)} = S^{(1)}$. For any $2 \leq i \leq N$, let*

$$\hat{S}^{(i)} = S^{(i)} \setminus \bigcup_{k=1}^{i-1} \hat{S}^{(k)}.$$

Let $\hat{N}$ be the number of non-empty sets $\hat{S}^{(i)}$.

Note that $\hat{N}$ may be smaller than N since there may be an element of ζ_N which is contained in the union of other elements of ζ_N. It is easy to verify by induction that the sets of $\zeta_{\hat{N}}$ are mutually exclusive and their union equals that of the original sets in ζ_N. We have the following:

Claim 1 $\left|V_{\mathcal{F}_+}(\zeta_N)\right| \le \left|V_{\mathcal{F}_+}(G(\zeta_N))\right|$, *where* $G()$ *denotes the output of procedure* G *applied to the argument.*

Proof: We make repetitive use of the following: let $A, B \subset X'$ be two non-empty sets and let $C = B \setminus A$. Then for any f, any $b \in \{0,1\}$, if $[e_A(f), e_B(f)] = [b, 0]$, then $[e_A(f), e_C(f)]$ may be either $[b, 0]$ or $[b, 1]$ since the elements in B which caused the product $e_B(f)$ to be zero may or may not also be in C. In the other case if $[e_A(f), e_B(f)] = [b, 1]$ then $[e_A(f), e_C(f)] = [b, 1]$. Hence

$$|\{[e_A(f), e_B(f)] : f \in \mathcal{F}_+\}| \le |\{[e_A(f), e_C(f)] : f \in \mathcal{F}_+\}| .$$

The same argument holds also for multiple $A_1, \ldots, A_k$, B and $C = B \setminus \bigcup_{i=1}^k A_i$. Let $\zeta_{\hat{N}} = G(\zeta_N)$. We now apply this to the following:

$$\begin{aligned}
&|\{[e_{S^{(1)}}(f), e_{S^{(2)}}(f), e_{S^{(3)}}(f), \ldots, e_{S^{(N)}}(f)] : f \in \mathcal{F}_+\}| \\
&= \left|\left\{\left[e_{\hat{S}^{(1)}}(f), e_{S^{(2)}}(f), e_{S^{(3)}}(f), \ldots, e_{S^{(N)}}(f)\right] : f \in \mathcal{F}_+\right\}\right| \quad (15)\\
&\le \left|\left\{\left[e_{\hat{S}^{(1)}}(f), e_{\hat{S}^{(2)}}(f), e_{S^{(3)}}(f), \ldots, e_{S^{(N)}}(f)\right] : f \in \mathcal{F}_+\right\}\right| \quad (16)\\
&\le \left|\left\{\left[e_{\hat{S}^{(1)}}(f), e_{\hat{S}^{(2)}}(f), e_{\hat{S}^{(3)}}(f), e_{S^{(4)}}(f) \ldots, e_{S^{(N)}}(f)\right] : f \in \mathcal{F}_+\right\}\right| \quad (17)\\
&\le \cdots \\
&\le \left|\left\{\left[e_{\hat{S}^{(1)}}(h), e_{\hat{S}^{(2)}}(h), e_{\hat{S}^{(3)}}(h), e_{\hat{S}^{(4)}}(h), \ldots, e_{\hat{S}^{(N)}}(h)\right] : f \in \mathcal{F}_+\right\}\right| \quad (18)
\end{aligned}$$

where (15) follows since using G we have $\hat{S}^{(1)} = S^{(1)}$, (16) follows by applying the above with $A = \hat{S}^{(1)}$, $B = S^{(2)}$ and $C = \hat{S}^{(2)}$, (17) follows by letting $A_1 = \hat{S}^{(1)}$, $A_2 = \hat{S}^{(2)}$, $B = S^{(3)}$, and $C = \hat{S}^{(3)}$. Finally, removing those sets $\hat{S}^{(i)}$ which are possibly empty leaves $\hat{N}$-dimensional vectors consisting only of the non-empty sets so (18) becomes $\left|\left\{\left[e_{\hat{S}^{(1)}}(f), \ldots, e_{\hat{S}^{(\hat{N})}}(f)\right] : f \in \mathcal{F}_+\right\}\right|$. □

Hence (11) is bounded from above as

$$\Gamma_{\mathcal{F}'_\gamma}(\zeta_N) \le \left|V_{\mathcal{F}_+}(G(\zeta_N))\right| . \quad (19)$$

Denote by $N^* = m - \ell + 1$ and define the following procedure which maps a generalized collection of sets in X to another.

Procedure Q: *Given a generalized collection* $\zeta_N = \{S^{(i)}\}_{i=1}^N$, $S^{(i)} \subset X$. *Construct* ζ_{N^*} *as follows: let* $Y = \bigcup_{i=2}^N S^{(i)}$ *and let the elements in* Y *be ordered according to their ordering on* X' *(we will refer to them as* y_1, y_2, $\ldots$*). Let* $S^{*(1)} = S^{(1)}$. *For* $2 \le i \le m - \ell + 1$, *let* $S^{*(i)} = \{y_{i-1}\}$.

We now have the following:

Claim 2 *For any* $\zeta_N \subset \mathbb{S}_\ell$ *with* $|\zeta_N| = m$, *we have*

$$\left|V_{\mathcal{F}_+}(G(\zeta_N))\right| \le \left|V_{\mathcal{F}_+}(Q(G(\zeta_N)))\right| .$$

Proof: Let $\zeta_{\tilde{N}} = Q(G(\zeta_N))$ and as before $\zeta_{\hat{N}} = G(\zeta_N)$. Note that by definition of Procedure Q, it follows that $\zeta_{\tilde{N}}$ consists of $\tilde{N} = N^*$ non-overlapping sets, the first $\tilde{S}^{(1)}$ having cardinality ℓ and $\tilde{S}^{(i)}$, $2 \le i \le \tilde{N}$, each having a single distinct element of X'. Their union satisfies $\bigcup_{i=1}^{\tilde{N}} \tilde{S}^{(i)} = X'$.

Consider the sets $V_{\mathcal{F}_+}(\zeta_{\hat{N}})$, $V_{\mathcal{F}_+}(\zeta_{\tilde{N}})$ and denote them simply by $\hat{V}$ and $\tilde{V}$. For any $\hat{v} \in \hat{V}$ consider the following subset of $\mathcal{F}_+$,

$$B(\hat{v}) = \{f \in \mathcal{F}_+ : \hat{v}(f) = \hat{v}\}.$$

We consider two types of $\hat{v} \in \hat{V}$. The first does *not* have the following property: there exist functions f_α, $f_\beta \in B(\hat{v})$ with $\theta^\gamma_{f_\alpha}(x) \neq \theta^\gamma_{f_\beta}(x)$ for at least one element $x \in X'$. Denote by $\theta^\gamma_f = [\theta^\gamma_f(x_1), \ldots, \theta^\gamma_f(x_m)]$. In this case all $f \in B(\hat{v})$ have the same $\theta^\gamma_f = \hat{\theta}$, where $\hat{\theta} \in \{0,1\}^m$. This implies that

$$e_{\tilde{S}^{(1)}}(f) = e_{\hat{S}^{(1)}}(f) = \hat{v}_1$$

while for $2 \leq j \leq \tilde{N}$ we have

$$e_{\tilde{S}^{(j)}}(f) = \hat{\theta}_{k(j)}$$

where $k : [N^*] \to [m]$ maps from the index of a (singleton) set $\tilde{S}^{(j)}$ to the index of an element of X' and $\hat{\theta}_{k(j)}$ denotes the $k(j)^{th}$ component of $\hat{\theta}$. Hence it follows that

$$|V_{B(\hat{v})}(\zeta_{\tilde{N}})| = |V_{B(\hat{v})}(\zeta_{\hat{N}})|.$$

Let the second type of $\hat{v}$ satisfy the complement condition, namely, there exist functions f_α, $f_\beta \in B(\hat{v})$ with $\theta^\gamma_{f_\alpha}(x) \neq \theta^\gamma_{f_\beta}(x)$ for at least one point $x \in X'$. If such x is an element of $\hat{S}^{(1)}$ then the first part of the argument above holds and we still have

$$|V_{B(\hat{v})}(\zeta_{\tilde{N}})| = |V_{B(\hat{v})}(\zeta_{\hat{N}})|.$$

If however there is also such an x in some set $\hat{S}^{(j)}$, $2 \leq j \leq \hat{N}$ then since the sets $\tilde{S}^{(i)}$, $2 \leq i \leq \tilde{N}$ are singletons then there exists some $\tilde{S}^{(i)} \subseteq \hat{S}^{(j)}$ with

$$e_{\tilde{S}^{(i)}}(f_\alpha) \neq e_{\tilde{S}^{(i)}}(f_\beta).$$

Hence for this second type of $\hat{v}$ we have

$$|V_{B(\hat{v})}(\zeta_{\tilde{N}})| \geq |V_{B(\hat{v})}(\zeta_{\hat{N}})|. \tag{20}$$

Together with the previous case, we have that (20) holds for any $\hat{v} \in \hat{V}$.

Now, consider any two distinct $\hat{v}_\alpha$, $\hat{v}_\beta \in \hat{V}$. Clearly, $B(\hat{v}_\alpha) \bigcap B(\hat{v}_\beta) = \emptyset$ since every f has a unique $\hat{v}(f)$. Moreover, for any $f_a \in B(\hat{v}_\alpha)$ and $f_b \in B(\hat{v}_\beta)$ we have $\tilde{v}(f_a) \neq \tilde{v}(f_b)$ for the following reason: there must exist some set $\hat{S}^{(i)}$ and a point $x \in \hat{S}^{(i)}$ such that $\theta^\gamma_{f_a}(x) \neq \theta^\gamma_{f_b}(x)$ (since $\hat{v}_\alpha \neq \hat{v}_\beta$). If $i = 1$ then they must differ on $\tilde{S}^{(1)}$, i.e., $e_{\tilde{S}^{(1)}}(f_\alpha) \neq e_{\tilde{S}^{(1)}}(f_\beta)$. If $2 \leq i \leq \hat{N}$, then such an x is in some set $\tilde{S}^{(j)} \subseteq \hat{S}^{(i)}$ where $2 \leq j \leq \tilde{N}$ and therefore $e_{\tilde{S}^{(j)}}(f_\alpha) \neq e_{\tilde{S}^{(j)}}(f_\beta)$. Hence no two distinct $\hat{v}_\alpha$, $\hat{v}_\beta$ map to the same $\tilde{v}$. We therefore have

$$\begin{aligned} \left|V_{\mathcal{F}_+}(\zeta_{\hat{N}})\right| &= \sum_{\hat{v} \in \hat{V}} |V_{B(\hat{v})}(\zeta_{\hat{N}})| \\ &\leq \sum_{\hat{v} \in \hat{V}} |V_{B(\hat{v})}(\zeta_{\tilde{N}})| \qquad (21) \\ &= |V_{\mathcal{F}_+}(\zeta_{\tilde{N}})| \end{aligned}$$

where (21) follows from (20) which proves the claim. □

Note that by construction of Procedure Q, the dimensionality of the elements of $V_{\mathcal{F}_+}(Q(G(\zeta_N)))$ is N^*, i.e., $m-\ell+1$, which holds for any ζ_N (even maximally overlapping) and X' as defined in (9) and (14). Let us denote by ζ_{N^*} any set obtained by applying Procedure G on any collection ζ_N followed by Procedure Q, i.e.,

$$\zeta_{N^*} = \left\{ S^{*(1)}, S^{*(2)}, \dots, S^{*(N^*)} \right\}$$

with a set $S^{*(1)} \subset X'$ of cardinality ℓ and

$$S^{*(k)} = \{x_{i_k}\}, \text{ where } x_{i_k} \in X' \setminus S^{*(1)}, \quad k = 2, \dots, N^*.$$

Hence we have

$$\begin{aligned} \max_{\zeta_N \subset \mathbb{S}_\ell, |\zeta_N|=m} \Gamma_{\mathcal{F}'_\gamma}(\zeta_N) &\leq \max_{\zeta_N \subset \mathbb{S}_\ell, |\zeta_N|=m} \left| V_{\mathcal{F}_+}\left(Q(G(\zeta_N))\right) \right| &(22)\\ &\leq \max_{\zeta_{N^*}:|\zeta_{N^*}|=m} \left| V_{\mathcal{F}_+}(\zeta_{N^*}) \right| &(23) \end{aligned}$$

where (22) follows from (11), (13) and Claims 1 and 2 while (23) follows by definition of ζ_{N^*}. Now,

$$\begin{aligned} \left| V_{\mathcal{F}_+}(\zeta_{N^*}) \right| &= \left| \{ [e_{S^{*(1)}}(f), \dots, e_{S^{*(N^*)}}(f)] : f \in \mathcal{F}_+ \} \right| \\ &\leq 2 \left| \{ [e_{S^{*(2)}}(f), \dots, e_{S^{*(N^*)}}(f)] : f \in \mathcal{F}_+ \} \right| &(24) \end{aligned}$$

where (24) follows trivially since $e_{S^{*(1)}}(f)$ is binary. So from (23) we have

$$\begin{aligned} \max_{\zeta_N \subset \mathbb{S}, |\zeta_N|=m} \Gamma_{\mathcal{F}'_\gamma}(\zeta_N) &\leq 2 \max_{\zeta_{N^*}:|\zeta_{N^*}|=m} \left| \{ [e_{S^{*(2)}}(f), \dots, e_{S^{*(N^*)}}(f)] : f \in \mathcal{F}_+ \} \right| \\ &\leq 2 \max_{x_1,\dots,x_{m-\ell} \in X} \left| \left\{ [\theta_f^\gamma(x_1), \dots, \theta_f^\gamma(x_{m-\ell})] : f \in \mathcal{F}_+ \right\} \right| &(25) \end{aligned}$$

where $x_1, \dots, x_{m-\ell}$ run over any $m-\ell$ points in X. Define the following infinite class of binary functions on X by

$$\Theta_{\mathcal{F}_+}^\gamma = \{ \theta_f^\gamma(x) : f \in \mathcal{F}_+ \}$$

and for any finite subset

$$X'' = \{x_1, \dots, x_{m-\ell}\} \subset X$$

let

$$\theta_f^\gamma(X'') = \left[\theta_f^\gamma(x_1), \dots, \theta_f^\gamma(x_{m-\ell}) \right]$$

and

$$\Theta_{\mathcal{F}_+}^\gamma(X'') = \{ \theta_f^\gamma(X'') : f \in \mathcal{F}_+ \}.$$

We proceed to bound $|\Theta_{\mathcal{F}_+}^\gamma(X'')|$.

The class $\Theta_{\mathcal{F}_+}^\gamma$ is in one-to-one correspondence with a class $\mathcal{C}_{\mathcal{F}_+}^\gamma$ of sets $C_f \subset X$ which are defined as

$$C_f = \{ x : \theta_f^\gamma(x) = 1 \}, \quad f \in \mathcal{F}_+.$$

We claim that any such set C_f equals the union of at most $K = \lfloor B/(2\gamma) \rfloor$ intervals. To see this, note that based on the general form of $f \in \mathcal{F}_+$ (see (6) and (7)) in order for $f(x) > \gamma$ for every x in an interval set $\mathfrak{I} \subset X$ then $\mathfrak{I}$ must be contained in an interval set of the form (5) and of length at least 2γ. Hence for any $f \in \mathcal{F}_+$ the corresponding set C_f is comprised of no more than K distinct intervals as $\mathfrak{I}$. Hence the class $\mathcal{C}^{\gamma}_{\mathcal{F}_+}$ is a subset of the class $\mathcal{C}_K$ of all sets that are comprised of the union of at most K subsets of X. A class H is said to *shatter* A if $\left|\{h_{|A} : h \in H\}\right| = 2^k$, where k is the cardinality of A. The Vapnik-Chervonenkis dimension of $\mathcal{H}$, denoted as $VC(\mathcal{H})$, is defined as the cardinality of the largest set shattered by $\mathcal{H}$. We claim that the VC-dimension of $\mathcal{C}_K$ is $\mathrm{VC}(\mathcal{C}_K) = 2K$. This can be shown by induction: it is clear that for $K = 1$, the class $\mathcal{C}_1$ can shatter any pair of points but cannot shatter a three point set since the alternating dichotomy $1, 0, 1$ cannot be obtained by a single interval set. Now, assume that it holds for $K - 1$, i.e., that $VC(\mathcal{C}_{K-1}) = 2(K - 1)$. Consider any set $E = \{x_1, x_2, x_3, \ldots, x_{2K}\} \subset X$ where $\{x_3, \ldots, x_{2K}\}$ is shattered by $\mathcal{C}_{K-1}$. The set of dichotomies of E obtained by the class $\mathcal{C}_K$ includes the set of dichotomies obtained by the product class $\mathcal{C}_1 \times \mathcal{C}_{K-1}$ on $\{\{x_1, x_2\}, \{x_3, \ldots, x_{2K}\}\}$. Since $\mathcal{C}_1$ shatters $\{x_1, x_2\}$ and by the inductive hypothesis $\mathcal{C}_{K-1}$ shatters $\{x_3, \ldots, x_{2K}\}$ then $VC(\mathcal{C}_K) \geq 2K$. To see that $VC(\mathcal{C}_K) \leq 2K$, note that it cannot obtain the alternating dichotomy $101\ldots01$ on any set $\{x_1, x_2, \ldots, x_{2K}, x_{2K+1}\}$ which proves the claim.

Continuing, it follows from the Sauer-Shelah lemma (see Sauer [1972]) that the growth of $\mathcal{C}^{\gamma}_{\mathcal{F}_+}$ on any finite set $X'' \subset X$ of cardinality $m - \ell$ (see (2)) satisfies

$$\mathit{\Gamma}_{\mathcal{C}^{\gamma}_{\mathcal{F}_+}}(X'') \leq \sum_{i=0}^{2K} \binom{m-\ell}{i}.$$

Since $|\mathit{\Theta}^{\gamma}_{\mathcal{F}_+}(X'')| = \mathit{\Gamma}_{\mathcal{C}^{\gamma}_{\mathcal{F}_+}}(X'')$ then from (8) and (25) it follows that

$$|\mathit{\Gamma}_{\mathcal{H}'_{\gamma}}(m)| \leq 2 \sum_{i=0}^{2\lfloor B/(2\gamma) \rfloor} \binom{m-\ell}{i}$$

which proves the statement of the theorem. □

Bibliography

M. Anthony and P. L. Bartlett. *Neural Network Learning:Theoretical Foundations.* Cambridge University Press, 1999.

M. Anthony and J. Ratsaby. Maximal width learning of binary functions. Technical Report LSE-CDAM-2006-11, (Centre for Discrete and Applicable Mathematics), Department of Mathematics, London School of Economics and Political Science, October 2006.

B. Bollobás. *Combinatorics: Set Systems, Hypergraphs, Families of vectors, and combinatorial probability.* Cambridge University Press, 1986.

J. Ratsaby. Information efficiency. In *The* 33^{rd} *International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM'07)*, volume LNCS 4362, pages 475–487. Springer-Verlag, 2007.

N. Sauer. On the density of families of sets. *J. Combinatorial Theory (A)*, 13:145–147, 1972.