Abstract

Consider a class $\mathcal{H}$ of binary functions $h : X \to \{-1,+1\}$ on a finite interval $X = [0,B] \subset \mathbb{R}$. Define the sample width of $h$ on a finite subset (a sample) $S \subset X$ as $\omega_S(h) \equiv \min_{x \in S} \omega_h(x)$ where $\omega_h(x) = h(x) \max \{a \geq 0 : h(z) = h(x), x-a \leq z \leq x+a\}$. Let $\mathcal{S}_\ell$ be the space of all samples in $X$ of cardinality $\ell$ and consider sets of wide samples, i.e., hypersets which are defined as $A_{\beta,h} = \{S \in \mathcal{S}_\ell : \omega_S(h) \geq \beta\}$. 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., the number of possible dichotomies obtained by intersecting all hypersets with a fixed collection of samples $S \in \mathcal{S}_\ell$ of cardinality $m$. The estimate is $2 \sum_{i=0}^{\lceil B/2\beta \rceil} (m-\ell)^i$.

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 ? we showed 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 (used also in ??) is based on the simple notion of width of $h$ at $x$ which is defined as

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

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

$$\omega_S(h) \equiv \min_{x \in S} |\omega_h(x)|.$$
This definition of width resembles the notion of sample margin of a real-valued function \( f \) (see for instance ?). 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 ?).

## 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 \( \ell \) be any fixed positive integer and define the space \( S_\ell \) of all samples \( S \subset X \) of size \( \ell \). On \( S_\ell \) consider sets of wide samples, i.e.,

\[
A_{\beta,h} = \{ S \in S_\ell : \omega_{S}(h) \geq \beta \}, \quad \beta > 0.
\]

We refer to such sets as **hypersets**. It will be convenient to associate with these sets the indicator functions on \( S_\ell \) which are denoted as

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

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

\[
\mathcal{H}'_{\gamma} = \{ h'_{\gamma,h} : h \in \mathcal{H} \}.
\]

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

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 \( \mathcal{H}'_{\gamma} \) hence one cannot simply measure its cardinality. Instead we apply a standard combinatorial measure of the complexity of a family of sets as follows: suppose \( Y \) is a general domain and \( \mathcal{G} \) is an infinite class of subsets of \( Y \). For any subset \( S = \{y_1, \ldots, y_n\} \subset Y \) let

\[
\Gamma_{\mathcal{G}}(S) = |\mathcal{G}|^{|S|} \quad (2)
\]

where \( \mathcal{G}|_S = \{\mathbb{I}_{G}(y_1), \ldots, \mathbb{I}_{G}(y_n) : G \in \mathcal{G}\} \). The *growth function* (see for instance ?) is defined as

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

It measures the rate in which the number of dichotomies obtained by intersecting subsets \( G \) of \( \mathcal{G} \) with a finite set \( S \) increases as a function of the cardinality \( n \) of \( S \) in the maximal case (it is also called the trace of \( \mathcal{G} \) in ?).

Since we are interested in hypersets as opposed to simple sets \( G \) (as above) then we consider the trace on a finite collection \( \zeta \subset S_\ell \) 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 \( \zeta \subset S_\ell \) let

\[
|\zeta| = \left| \bigcup_{S : S \in \zeta} S \right| \quad (3)
\]
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.

\[
\Gamma_{\mathcal{H}_\gamma}(m) = \max_{\zeta \subseteq S_\ell, |\zeta| = m} \Gamma_{\mathcal{H}_\gamma}(\zeta).
\]

### 3 Main result

Let us state the main result of the paper.

**Theorem 1** Let \( \ell, m > 0 \) be finite integers and \( B > 0 \) a finite real number. Let \( \mathcal{H} \) be the class of binary functions on \([0, B]\) (with only simple discontinuities). For a given width parameter value \( \gamma > 0 \), the corresponding hyperclass \( \mathcal{H}_\gamma \) on the space \( S_\ell \) has a growth which is bounded as

\[
\Gamma_{\mathcal{H}_\gamma}(m) \leq 2 \left( \frac{2[2^{B/(2\gamma)}]}{(m - \ell)} \right).
\]

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

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

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 \( I_R(x) \) to represent the indicator function of the statement \( x \in R \). In case of an interval set \( R = \langle a, b \rangle \) we write \( I_{\langle a, b \rangle} \).

**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 is as simple as we can find. 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 \( \{R_i\}_{i=1,2,...} \) which satisfies the following properties: (a) \([0, B] = \bigcup_{i=1,2,...} R_i \) 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

\[
h(x) = \pm \sum_{i=1,2,...} (-1)^i I_{R_i}(x).
\]

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

\[
R_i = (a_{i-1}, a_i), \quad i = 1, 2, \ldots \tag{5}
\]

with the first left end point being \( a_0 = 0 \). Note that different choices for \( \langle a, b \rangle \) (see earlier definition of a generalized interval \( \langle a, b \rangle \)) give different sets \( R_i \) and hence different functions \( h \). For instance, suppose \( X = [0, 7] \) 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

\[
\mu_i = \frac{a_i + a_{i+1}}{2}, \quad i = 1, 2, \ldots
\]

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

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

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 \( a_i, i = 1, 2, \ldots \) the function \( f \) is invariant to the type of intervals \( R_i = (a_{i-1}, a_i) \) 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

![Figure 1: \( h \) (solid) and its corresponding \( f \) (dashed) on \( X = [0, B] \) with \( B = 800 \)](image-url)
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
\[
F_+ = \{|f| : f \in F\}.
\]
(7)
It follows that the hyperclass \( \mathcal{H}_\gamma' \) may be represented in terms of the class \( F_+ \) as follows: define the hypersets
\[
A_{\beta,f} = \{ S \in \mathcal{S}_\ell : f(x) \geq \beta, x \in S \}, \quad \beta > 0, f \in F_+
\]
with corresponding hyperconcepts \( f'_{\gamma,f} = \mathbb{I}_{A_{\beta,f}}(S) \), let
\[
F_\gamma' = \{ f'_{\gamma,f} : f \in F_+ \}
\]
and
\[
\mathcal{H}_\gamma' = F_\gamma'.
\]
(8)
Hence, it suffices to compute the growth function \( \Gamma_{F_\gamma'}(m) \).

Let us now begin to analyze the hyperclass \( F_\gamma' \). By definition, \( F_\gamma' \) is a class of indicator functions of subsets of \( \mathcal{S}_\ell \). Denote by \( \zeta_N \subset \mathcal{S}_\ell \) 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 \( \zeta_N \) such that \(|\zeta_N| = m \)
(9)
where recall the definition of cardinality is according to (3). Let us denote the individual components of \( \zeta_N \) by \( S^{(j)} \in \mathcal{S}_\ell, 1 \leq j \leq N \) hence
\[
\zeta_N = \{ S^{(1)}, \ldots, S^{(N)} \}.
\]
The growth function may be expressed as
\[
\Gamma_{F_\gamma'}(m) = \max_{\zeta_N \subset \mathcal{S}_\ell, |\zeta_N| = m} \Gamma_{F_\gamma'}(\zeta_N) \equiv \max_{\zeta_N \subset \mathcal{S}_\ell, |\zeta_N| = m} \left| \left\{ [f'(S^{(1)}), \ldots, f'(S^{(N)})] : f' \in F_\gamma' \right\} \right|.
\]
(10)
Denote by \( S^{(j)}_i \) 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
\[
\Gamma_{F_\gamma'}(\zeta_N)
= \left| \left\{ \left[ \mathbb{I} \left( \min_{x \in S^{(1)}} f(x) > \gamma \right), \ldots, \mathbb{I} \left( \min_{x \in S^{(N)}} f(x) > \gamma \right) \right] : f \in F_+ \right\} \right|
= \left| \left\{ \prod_{j=1}^{\ell} \mathbb{I} \left( f(S^{(1)}_j) > \gamma \right), \ldots, \prod_{j=1}^{\ell} \mathbb{I} \left( f(S^{(N)}_j) > \gamma \right) : f \in F_+ \right\} \right|.
\]
(11)
Order the elements in each component of \( \zeta_N \) by the underlying ordering on \( X \). Then put the sets in lexical ordering starting with the first up to the \( \ell \)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
\[ \theta f^j(x) \equiv \mathbb{I}(f(x) > \gamma) \]  \hspace{1cm} (12)

(we will sometimes write \( \theta_f(x) \) for short). For any sample \( S^{(i)} \) of cardinality \( |S^{(i)}| \geq 1 \) let
\[ e_{S^{(i)}}(f) = \prod_{j=1}^{|S^{(i)}|} \theta_f(S^{(i)}_j). \]

Then for \( \zeta_N \) we denote by
\[ v_{\zeta_N}(f) \equiv [e_{S^{(1)}}(f), \ldots, e_{S^{(N)}}(f)] \]
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
\[ \Gamma_{\mathcal{F}^+}(\zeta_N) = |V_{\mathcal{F}^+}(\zeta_N)|. \]  \hspace{1cm} (13)

Denote by \( X' \) the union
\[ \bigcup_{j=1}^N S^{(j)} = X' = \{x_i\}_{i=1}^m \subset X \]  \hspace{1cm} (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 \( \zeta_N \) is left implicit. We will need the following procedure which maps \( \zeta_N \) to a generalized collection.

**Procedure G**: Given \( \zeta_N \) 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 \( \zeta_N \) which is contained in the union of other elements of \( \zeta_N \). 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 \( \zeta_N \). We have the following:

**Claim 1** \( |V_{\mathcal{F}^+}(\zeta_N)| \leq |V_{\mathcal{F}^+}(G(\zeta_N))| \).
For any Claim 2

We now have the following:

$$X$$ satisfies $$\ell$$ follows: let $$Q$$

Proof: Let $$\hat{G}$$ be the generic collection of sets in $$A$$.

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

$$|\{|e_{S^{(1)}}, e_{S^{(2)}}, e_{S^{(3)}}, \ldots, e_{S^{(N)}} : f \in \mathcal{F}_+|\} = |\{|e_{S^{(1)}}, e_{S^{(2)}}, e_{S^{(3)}}, \ldots, e_{S^{(N)}} : f \in \mathcal{F}_+|\} = |\{|e_{S^{(1)}}, e_{S^{(2)}}, e_{S^{(3)}}, \ldots, e_{S^{(N)}} : f \in \mathcal{F}_+|\} = 15$$

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 $$A_1 = \hat{S}^{(1)}$$, $$A_2 = S^{(2)}$$, $$B = S^{(3)}$$, 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

$$\Gamma_G(\hat{\zeta}) \leq |V_{\mathcal{F}_+}(G(\hat{\zeta}))|.$$  

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 $$\zeta_N = \{S^{(i)}\}_{i=1}^N \subset X$$. Construct $$\zeta_{N^*}$$ 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 $$2 \leq i \leq m - \ell + 1$$, let $$S^{(i)} = \{y_{i-1}\}$$.

We now have the following:

**Claim 2** For any $$\zeta_N \subset S_\ell$$ with $$|\zeta_N| = m$$, then

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

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

7
Consider the sets $V_{\mathcal{F}_+}(\zeta_N)$, $V_{\mathcal{F}_+}(\zeta_N)$ and denote them simply by $\hat{V}$ and $\hat{\hat{V}}$. 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 $f_\alpha, f_\beta \in B(\hat{v})$ with $\theta_f^\alpha(x) \neq \theta_f^\beta(x)$ for at least one element $x \in X'$. Denote by $\theta_f^\gamma \equiv [\theta_f^\gamma(x_1), \ldots, \theta_f^\gamma(x_m)]$. Then in this case all $f \in B(\hat{v})$ have the same $\theta_f^\gamma = \hat{\theta}$, where $\hat{\theta} \in \{0, 1\}^m$. This implies that

$$
eq \hat{v}_1$$

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

$$
eq \hat{v}_k(j)$$

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

$$
|\{B(\hat{v})\}(\zeta_N)| = |\{B(\hat{v})\}(\zeta_N)|.
$$

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

$$
|\{B(\hat{v})\}(\zeta_N)| = |\{B(\hat{v})\}(\zeta_N)|.
$$

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

$$
eq \hat{v}_\alpha.$$ 

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

$$
|\{B(\hat{v})\}(\zeta_N)| \geq |\{B(\hat{v})\}(\zeta_N)|.
$$

Combining the above, then (20) holds for any $\hat{v} \in \hat{V}$.

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

$$
|\{V_{\mathcal{F}_+}(\zeta_N)| = \sum_{\hat{v} \in \hat{V}} |\{B(\hat{v})\}(\zeta_N)| 
\leq \sum_{\hat{v} \in \hat{V}} |\{B(\hat{v})\}(\zeta_N)| 
= |\{V_{\mathcal{F}_+}(\zeta_N)|
$$

8

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

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 \(\zeta_N\) (even maximally overlapping) and \(X'\) as defined in (9) and (14). Let us denote by \(\zeta_{N^*}\) any set obtained by applying Procedure \(G\) on any collection \(\zeta_N\) followed by Procedure \(Q\), i.e.,

\[
\zeta_{N^*} = \left\{ S^{(1)}, S^{(2)}, \ldots, S^{(N^*)} \right\}
\]

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

\[
S^{(k)} = \{x_{ik}\}, \text{ where } x_{ik} \in X' \setminus S^{(1)}, \quad k = 2, \ldots, N^*.
\]

Hence we have

\[
\max_{\zeta_N \subseteq S, |\zeta_N| = m} \Gamma_{\mathcal{F}^+}(\zeta_N) \leq \max_{\zeta_{N^*} \subseteq S, |\zeta_{N^*}| = m} |V_{\mathcal{F}^+}(Q(G(\zeta_N)))| \leq \max_{\zeta_{N^*} \subseteq S, |\zeta_{N^*}| = m} |V_{\mathcal{F}^+}(\zeta_{N^*})| \tag{22}
\]

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

\[
|V_{\mathcal{F}^+}(\zeta_{N^*})| = |\{[e_{S^{(1)}(f)}, \ldots, e_{S^{(N^*)}}(f)] : f \in \mathcal{F}_+\}| \leq 2 |\{[e_{S^{(2)}(f)}, \ldots, e_{S^{(N^*)}}(f)] : f \in \mathcal{F}_+\}| \tag{24}
\]

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

\[
\max_{\zeta_N \subseteq S, |\zeta_N| = m} \Gamma_{\mathcal{F}^+}(\zeta_N) \leq 2 \max_{\zeta_{N^*} \subseteq S, |\zeta_{N^*}| = m} \left| \left\{ [\theta^+_f(x_1), \ldots, \theta^+_f(x_{m-\ell})] : f \in \mathcal{F}_+ \right\} \right| \tag{25}
\]

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

\[
\Theta^+_{\mathcal{F}_+} = \{ \theta^+_f(x) : f \in \mathcal{F}_+ \}
\]

and for any finite subset

\[
X'' = \{x_1, \ldots, x_{m-\ell}\} \subset X
\]

let

\[
\theta^+_f(X'') = [\theta^+_f(x_1), \ldots, \theta^+_f(x_{m-\ell})]
\]

and

\[
\Theta^+_{\mathcal{F}_+}(X'') = \{ \theta^+_f(X'') : f \in \mathcal{F}_+ \}.
\]

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

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

\[
C_f = \{ x : \theta^+_f(x) = 1 \}, \quad f \in \mathcal{F}_+.
\]
We claim that any such set $C_f$ equals the union of at most $K \equiv \lfloor B/(2\gamma) \rfloor$ intervals. To see this, note that based on the general form of $f \in F_+$ (see (6) and (7)) in order for $f(x) > \gamma$ for every $x$ in an interval set $\mathcal{I} \subset X$ then $\mathcal{I}$ must be contained in an interval set of the form (5) and of length at least $2\gamma$. Hence for any $f \in F_+$ the corresponding set $C_f$ is comprised of no more than $K$ distinct intervals as $\mathcal{I}$. Hence the class $C^f_{\mathcal{F}_+}$ is a subset of the class $C_K$ of all sets that are comprised of the union of at most $K$ subsets of $X$. A class $H$ is said to \emph{shatter} $A$ if $|\{h \upharpoonright A : h \in H\}| = 2^k$. The Vapnik-Chervonenkis dimension of $H$, denoted as $VC(H)$, is defined as the cardinality of the largest set shattered by $\mathcal{H}$. It is easy to show that the VC-dimension of $C_K$ is $VC(C_K) = 2K$. Hence it follows from the Sauer-Shelah lemma (see ?) that the growth of $C^f_{\mathcal{F}_+}$ on any finite set $X'' \subset X$ of cardinality $m - \ell$ (see (2)) satisfies

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

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

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

which proves the statement of the theorem.\hfill \square