New Bounds and Variants of VC Dimension of Boolean Function Classes

A THESIS

submitted by

AMIT ROY

for the award of the degree

of

MASTER OF SCIENCE

(by Research)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

June 2021

THESIS CERTIFICATE

This is to certify that the thesis titled **New Bounds and Variants of VC Dimension of Boolean Function Classes**, submitted by **Amit Roy**, to the Indian Institute of Technology Madras, for the award of the degree of **Master of Science (by Research)**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Jayalal Sarma
Research Guide
Professor
Dept. of Computer Science &
Engineering
IIT Madras, Chennai - 600 036

Place: Chennai

Date:

LIST OF PAPERS BASED ON THESIS

1. Amit Roy and Jayalal Sarma. On Alternation, VC-dimension and k-fold Union of Sets. Proceedings in 11th *European Conference on Combinatorics, Graph Theory and Applications* (EUROCOMB 2021), Sep 6-10, 2021, Barcelona

ACKNOWLEDGEMENTS

And the day came, when the risk to remain tight in a bud
was more painful than the risk it took to blossom...

Anaïs Nin

I want to thank my advisor, Jayalal Sarma for the constant guidance and encouragement that he has imparted to me since the day of joining this wonderful institute. His mentorship was not just limited to academics but also emotionally in such trying times. He is so accessible that at times I have felt that he was never offline. His curiosity, compassion, the excitement of problem-solving, teaching techniques is something which I always look up to.

The lab was the place where I spent most of my time in this institute. I want to thank all the members of the lab: Ankit, Anoop, Barath, Dinesh, Janani, Mansi, Neha, Omprakash, Purnata, Ramya, Sagar, Sampriti, Sreyas, Sumathi, and Swaroop who contributed to the fostering environment. I want to thank members from the TCS lab: Arasu, Amit, Anil, Govind, Nisha, and Santhini. I want to thank the members of the GTC: Prof. Balaraman Ravindran, Dr. Yadu Vasudev, and Dr. Jayanthan A V for their valuable feedbacks and generous time.

The time here flew in a blink of an eye and it was due to wonderful moments that I had with all my friends. They all have become very dear. I want to thank: Prithwish, Bhagyashree, Diptanshu, Rahul, Gaurav, Madhura, Aakriti, Sadbhavana, Sampriti, Divyanshu, Dipayan, Saikat, Saikumar, Jayesh, Radhakrishna. We enjoyed our life to the fullest whether it was playing mafia, dumb charades, Pictionary on birthday celebrations, late-night discussions on life and career at Ramu tea, or dining in every corner of the city or trips to Pondicherry, Ooty, or early morning Bessy beaches. The group was the family who helped through thick and thin. I have missed out a lot of names but they are always close to my heart. I can never forget all my squad from Cricket, Football, Volleyball, Frisbee, Swimming, Marathon, Badminton, and Cycling.

I want to thank Prof. V Kamakoti, who funded my entire MS period generously through his ISEA project. I also want to thank Prof. C Pandurangan, with whom I fondly remember our first warm interaction in CCD where he gave a glimpse about life, history, and art of southern India. I still remember one of his remarks, *“ I am older than some of the trees here"* ! I am also grateful to Yadu Vasudev, Raghavendra Rao, Meghana Nasre, and Narayanaswamy NS for their wonderful courses and all the interactions that we had. They have a great part to play in my thinking and research. I always look upon them for inspiration. I am grateful to my parents for always standing by my side. Finally, I want to thank the Computer Science Department and this wonderful institute that gave me the chance to be a part of this place. I wish that all my countrymen had the privilege to go to an institution such as this.

ABSTRACT

KEYWORDS: VC Dimension, Boolean Functions, Extremal Sets, k-union, Alternation, Certificate Complexity

We study Vapnik-Chervonenkis (VC) dimension of families of Boolean functions, and also for a Single Boolean function. Vapnik-Chervonenkis Dimension (VC-dimension) is a combinatorial measure of a set system of subsets of a universe, developed by Vapnik and Chervonenkis during 1960-1990 and has found deep applications in the area of statistical learning theory, discrete, and computational geometry. Let U be the universe and $\mathcal{F}$ be a set of subsets of U. The family $\mathcal{F}$ is said to *shatter* an S subset of U if for all subset $S' \subseteq S$, there is an $F \in \mathcal{F}$ such that $S \cap F = S'$. VC-dimension of a family is the size of the largest set shattered by the family. VC-dimension bounds of families give a general lower and upper bound on sample complexity to learn the concepts from the hypothesis class in the Probability Approximately Correct (PAC) model.

VC-dimension of Monotone family is already known to be $\binom{n}{n/2}$ due to Procaccia and Rosenschein (2006). Thus it is natural to study the family of non-monotone Boolean functions. *Alternation* of a Boolean function is a measure of non-monotonicity and we parameterize a family of Boolean functions with maximum *alternation* k and prove asymptotically matching lower and upper bound of VC-dimension for the family. For a special case of $k = 1$, the family contains *Monotone* functions and its negation for which we give exact bounds of VC-dimension . As an application, we show tightness of VC-dimension bounds for k-fold union, by explicitly constructing a family $\mathcal{F}$ of subsets of $\{0,1\}^n$ such that k-fold union of the family, $\mathcal{F}^{k\cup} = \left\{ \bigcup_{i=1}^{k} F_i \mid F_i \in \mathcal{F} \right\}$ must have VC-dimension at least $\Omega(dk)$ and that this bound holds even when the union is over disjoint sets from $\mathcal{F}$. This provides a non-geometric set system achieving this bound. In addition we also show *shattering extremal* properties of the *Monotone* family.

On the characteristic vectors variant of VC-dimension of a Boolean function, we compute VC-dimension bounds for functions such as AND, OR, Threshold, Parity, Majority, Monotone, and Symmetric. We also gave a characterization to put this problem

into Σ_3^p and showed that the problem is NP-hard. As the relation between various complexity measures of a Boolean function is central research of interest, we show a connection of VC-dimension with *certificate complexity* and thus establishing a relation with other measures. We show that $\mathsf{VC}(f) + C(f) \geq n$ for any Boolean function f.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

bs	Block Sensitivity
C	Certificate Complexity
NP	Complexity Class NP
P	Complexity Class P
s	Sensitivity
SVC	Subspace Vapnik-Chervonenkis
s-extremal	Shattering-Extremal
VC	Vapnik-Chervonenkis

CHAPTER 1

Introduction

Computational Complexity Theory is a subfield of Theoretical Computer Science that aims toward classifying problems according to their difficulty or inherent complexity. There are several models of computation such as *Turing Model, Post systems, μ-recursive functions, λ-calculus, and combinatory logic* which have their own power and limitations but computationally equivalent. The complexity used for classifying may depend on models being used such as time and space complexity for one, or randomness, alternation, circuit size for another model. One of the most general models of computation is the Turing machine model and the complexity measures attached with the model are Time and Space. Considering time as a resource, problems can be classified on the amount of time (relative to the input size) taken to find a solution to the problem. Similarly, problems can be classified on the amount of space being used to solve the problem. A language L is a set of strings ($L \subseteq \Sigma^*$ where Σ is set of symbols and Σ^* denotes set of all finite strings over Σ). Any computational problem can be formally thought of as a language. A Turing Machine is said to solve the problem if and only if it accepts the language L corresponding to the problem. We can encode a problem into a set of strings by representing various objects and operations as strings of some fixed alphabet Σ. For any language L, let us define $L_n, n \geq 0$ as $L \cap \Sigma^n$ i.e. all strings of length n. There is a natural way to associate a family of Boolean functions for a language $L \subseteq \Sigma^*$. The family of Boolean functions are the characteristic function $\{\mathbb{1}_{L_n}\}$ where $\mathbb{1}_{L_n}(x) = 1 \iff x \in L_n$.

Another model of computation based on the above discussion is Boolean circuit model. A Boolean circuit is a directed acyclic graph where nodes are called gates (*AND, OR, NOT*) and leaves (*in-degree*$= 0$) are variables. For any Boolean function $f : \{0,1\}^n \to \{0,1\}$, there exists a Boolean circuit that computes the function. Strong lower bounds were developed in the early 80s for this model which turned out to be a promising approach towards central questions in Theoretical Computer Science. A circuit family $\mathcal{C} = \{C_i\}_{i \in \mathbb{N}}$ is an infinite set of circuits such that C_i has i input gates. A circuit family defines a function $f : \{0,1\}^* \to \{0,1\}$ in the most natural way as

$f(x) = C_{|x|}(x)$. The size and depth of a circuit family is measured as a function of number of input gates denoted as $SIZE(C)$ and $DEPTH(C)$. The total number of gates in the circuit is the size of the Boolean circuit and length of the longest path from the leaves to the root is called the depth of the circuit. The following complexity classes can be defined based on these two measures. A language L is said to be in $SIZE(t)$ iff there exists a circuit family $\mathcal{C} = \{C_i\}$ such that $SIZE(C_i) = O(t(i))$. A language L is said to be in $DEPTH(t)$ iff there exists a circuit family $\mathcal{C} = \{C_i\}$ such that $DEPTH(C_i) = O(t(i))$.

There are several well-studied measures of complexity of a Boolean function such as *degree, sensitivity, block-sensitivity, certificate complexity* etc. Sensitivity and block sensitivity were first studied by Cook *et al.* (1986) for certain parallel RAM algorithms but later it was found to be related to many other parameters such as *degree, certificate complexity*. One major research direction has been to understand the relation between all these measures. These measures can also be used to separate Boolean circuit classes. To show that a Boolean function family $\{f_n\}_{n\geq 1}$ cannot be computed by a certain class of Boolean circuits, one needs to identify a complexity measure on Boolean functions such that the measure of the function f_n is very different from the measure on every function computable by the class of circuits.

1.1 VC-dimension and PAC Learnability

Vapnik-Chervonenkis Dimension (VC-dimension) is a combinatorial measure of a set system of subsets of a universe, developed by Vapnik and Chervonenkis during 1960-1990 (Vapnik and Chervonenkis (1971)) and has found deep applications in the area of statistical learning theory, discrete and computational geometry. Let U be universe and $\mathcal{F}$ be a set of subsets of U. The family $\mathcal{F}$ is said to *shatter* a subset $S \subseteq U$ if for all subset $S' \subseteq S$, there is an $F \in \mathcal{F}$ such that $S \cap F = S'$. VC-dimension is the size of the largest set shattered by the family.

In the case when $U = \{0,1\}^n$, each $F \in \mathcal{F}$ can be interpreted as the positive inputs of a Boolean function. That is, as the set $f^{-1}(1) = \{x \in \{0,1\}^n \mid f(x) = 1\}$ of some Boolean function $f : \{0,1\}^n \to \{0,1\}$. Hence, any family $\mathcal{F}$ can be equivalently interpreted as a family of Boolean functions. Under this interpretation, the VC-dimension of a family is directly related to the complexity of learning the Boolean functions in the family in the Probably Approximately Correct (PAC) model by Valiant (1984). This connection is known to yield matching upper (Blumer *et al.* (1989)) and lower (Ehrenfeucht *et al.* (1989)) bounds for the number of samples required to learn the functions from the family. Motivated by this, there have been several works exploring tight VC dimensions bounds for various families of Boolean functions with n variables. Tight bounds are known for VC-dimension of various subfamilies of Boolean functions such as Boolean terms (conjunction of literals) - n (Natschläger and Schmitt (1996)), k-Decision Lists - n^k (Rivest (1987)), Symmetric Functions - $n+1$ (Ehrenfeucht *et al.* (1989)).

An important class of functions for which a characterization is known is the class of monotone functions (Procaccia and Rosenschein (2006)) where the VC-dimension was established to be exactly $\binom{n}{n/2}$. A subclass of this family - namely the monotone terms was also studied in Natschläger and Schmitt (1996) to establish a tight linear bound for VC-dimension.

There are several other applications of VC-dimension apart from sample complexity bounds of learning algorithms. Koiran (1996) uses VC-dimension to show a $\Omega(n^{\frac{1}{4}})$ lower bound on size of sigmoidal circuit computing a specific function in AC_2^0. Kremer *et al.* (1995) gives a lower bound on *One-Way Randomized Communication Complexity* of function using VC-dimension of the function. In another work, Klauck (2000) showed that one-way quantum communication complexity of a function f is lower bounded by VC-dimension of f. Bhattacharya *et al.* (2020) shows a new lower bound for communication complexity of the Disjointness problem using VC-dimension . Thus exposure of VC-dimension is not limited to sample complexity bounds. The problem of computing VC-dimension of a family is a computationally difficult problem and showed to be a Σ_3^p-complete problem (Schafer (1996)). In this thesis, we also study VC-dimension of a Boolean function as a Boolean function complexity measure and try to establish a connection with other Boolean function measures.

1.2 Thesis Contribution

Due to the connection of VC-dimension and sample complexity lower bounds, we take the work further to compute VC-dimension bounds for family of non-monotone Boolean functions. We study VC-dimension of Boolean function families parameterized by the *alternation* and show the following results:

Exact VC-dimension for family of functions with alternation 1: We show that family of functions with alternation 1 has VC-dimension exactly $\binom{n}{n/2} + 1$. (Theorem 4.1.4).

Tight bounds for VC-dimension for family of functions with alternation k: We show (Theorem 4.2.4 and 4.2.5) that the family of functions $\mathcal{F}$ with alternation k has VC-dimension satisfying : $\sum_{i=\frac{n-k}{2}}^{\frac{n+k}{2}} \binom{n}{i} \leq \mathsf{VC}(\mathcal{F}_k) \leq O\left(k \times \binom{n}{n/2}\right)$. For $k \leq \sqrt{n}$ the upper and lower bounds are asymptotically of the same order and hence the bound is tight in general.

Application to VC-dimension of disjoint union of families: If a family $\mathcal{F}$ is of VC-dimension $\leq d$, how large can the VC-dimension of the k-fold union family, defined as, $\mathcal{F}^k = \{\cup_{i=1}^{k} A_i \mid A_i \in \forall i, A_i \in \mathcal{F}\}$ be? Blumer *et al* Blais *et al.* (2015) and Haussler and Welzl Haussler and Welzl (1987) showed that the VC-dimension is at most $O(dk \log k)$. This bound was shown to be tight by Eisenstat and Angluin (2007) who shows existence of a geometric family with VC-dimension at most d and the k-fold union has VC-dimension at least $\Omega(dk \log k)$. The family constructed were point sets in a plane. Using our methods, we construct a family of Boolean functions such that k-fold union has VC-dimension at least $\Omega(dk)$ (Theorem 4.3.4) even when the unions in the k-fold union are restricted to k-fold disjoint union.

s-extremal Families: Let $Sh(\mathcal{F})$ be set of sets shattered by family $\mathcal{F}$. For an *s-extremal* family, $|Sh(\mathcal{F})| = |\mathcal{F}|$. We show that family of *Monotone* Boolean functions is *s-extremal* (Proposition 4.4.1). We also demonstrated a function g such that upon removing from the family of *Monotone* Boolean functions, resultant family is still *s-extremal*. Thus showing that *Meszaros-Ronyai* Conjecture holds true for family of *Monotone* Boolean functions (Proposition 4.4.4).

Subspace VC Dimension or SVC: We study Subspace VC-dimension motivated due to a possible role in separating the classes AC^0 and CC^0. We first observe that for any family $\mathcal{F}$, $\mathsf{SVC}(\mathcal{F}) \leq \mathsf{VC}(\mathcal{F})$. We computed SVC of *Monotone, Family of Parity, Symmetric* Boolean families.

We study VC-dimension of a single Boolean function, determine the hardness of computation, and establish few connections with other Boolean function complexity measures.

VC-dimension of Boolean function We compute VC-dimension of Boolean functions such as *Parity, Majority, Threshold*(Th_k^n)*, Monotone, k-slice, Symmetric* function.

Complexity of Computing On the Complexity of Computing $\mathsf{VC}(f)$, we proved that this problem is NP-hard (Proposition 5.3.3) and gave a characterization which puts the problem into Σ_3^p (Proposition 5.3.2).

Relation with other complexity measures We gave a relation between *certificate* complexity and $\mathsf{VC}(f)$ which establishes path with other complexity measures of a Boolean function (Proposition 5.2.4).

A part of these results appeared in Roy and Sarma (2021).

1.3 Thesis Outline

In this chapter, we gave a brief introduction to the problem and the area. The rest of the thesis is organized as follows: We explain the background and preliminaries needed in Chapter 2. In Chapter 3, we give a broad view of the problems in Computational Complexity, known VC-dimension bounds of families and recent application of VC-dimension . In Chapter 4 and 5, we discuss the main contribution of the thesis in detail and provide full proofs of all technical statements. Finally, we conclude the thesis in Chapter 6 with directions for future work.

CHAPTER 2

Preliminaries

In this chapter, we define all the variants of VC-dimension . We will also see, all the complexity measure associated with a Boolean function which has appeared in the thesis.

We use the notation $[n] = \{1, 2, \ldots, n\}$. For any set U, 2^U denotes the power set of U. All $\log$ are to the base 2 unless stated otherwise. In this thesis, whenever we talk about family of Boolean functions, our universe is $U = \{0, 1\}^n$. When the family is a single Boolean function, our universe is $U = \{1, 2, \ldots, n\}$. Let $x \in \{0, 1\}^n$ then x_i denotes the i^{th} bit of x. For any $x \in \{0, 1\}^n$, $x^{\oplus i}$ denotes the i^{th} bit flipped i.e. 0 becomes 1 and vice-versa. For $x, y \in \{0, 1\}^n$, $x \oplus y$ denotes bit-wise XOR-operation of x, y and same holds for other Boolean operators. Weight of the input x is defined as number of 1's in the input x and denoted by $wt(x)$. We use traditional notation for set operations: $\cup$ for union, $\cap$ for intersection and $\triangle$ for symmetric difference of two sets. In this thesis, we refer to family and concept class interchangeably.

2.1 Vapnik-Chervonenkis Dimension

In this section, we define all the variants of VC-dimension . Let U be a universe and $\mathcal{F}$ be a set of subsets of U. The family $\mathcal{F}$ is said to shatter a subset $S \subseteq U$ if for all subset $S' \subseteq S$, there is an $F \in \mathcal{F}$ such that $S \cap F = S'$. Notice that it is easy to shatter small sets, especially the empty set, and U is not shattered unless $\mathcal{F} = \mathcal{P}(U)$. The VC-dimension of $\mathcal{F}$ is said to be the largest d such that there is a set S of size d shattered by the family.

Definition 2.1.1 (**VC-dimension of a Boolean function family**). Let $\mathcal{F}$ be a collection of Boolean functions $f : \{0, 1\}^n \to \{0, 1\}$. Interpret each $f \in \mathcal{F}$ as a set of strings $\{x \in \{0, 1\}^n \mid f(x) = 1\}$. Thus, $\mathcal{F}$ can also be interpreted as a set of subsets of $U = \{0, 1\}^n$. The VC-dimension of the family of Boolean functions is defined to be the size of the largest set $S \subseteq U$ shattered by the family $\mathcal{F}$.

Definition 2.1.2 (VC-dimension of a Boolean function)**.** Let $U = \{1, 2, \dots n\}$. For a subset $F \subseteq U$, denote $x_F \in \{0,1\}^n$ as the n-bit string for which i^{th} bit is 1 if $i \in F$ and 0 otherwise. For a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, associate a set of subsets:

$$\mathcal{F} = \{F \subseteq U \mid f(x_F) = 1\}$$

The VC-dimension of a Boolean function f, denoted by $\mathsf{VC}(f)$, is the size of the largest set $S \subseteq [n]$ shattered by the family. Observe that $\mathcal{F} \subseteq 2^{[n]}$.

For clarity, we consider the following Boolean function. Consider $f(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee (x_2 \wedge x_3) \vee (x_3 \wedge x_1)$. Suppose the universe to be $U = \{1, 2, 3\}$. As per the above definition,

$$\mathcal{F} = \{\{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$$

For this example, the VC-dimension is 1 since we can shatter the sets $\{1\}, \{2\}, \{3\}$, and we cannot shatter any set greater than size 1.

For a given circuit C which computes a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, VC-dimension can also be defined as given below. This was studied by Schafer (1996).

Definition 2.1.3 (Schafer (1996))**.** Let f be a Boolean function from $\{0,1\}^m \rightarrow \{0,1\}$. Let $m = n + \ell$ and $U = \{0,1\}^n$. The family $\mathcal{F}$ is of size 2^ℓ and is indexed by $\{0,1\}^\ell$. Formally,

$$\mathcal{F} = \{F_w \mid w \in \{0,1\}^\ell\} \text{ where } F_w = \{x \in \{0,1\}^n \mid f(wx) = 1\}$$

where wx denotes concatenation of two strings w and x. VC-dimension of a Boolean function f is the size of the largest set $S \subseteq U$ shattered by the family $\mathcal{F}$.

We define now another variant called Subspace VC-dimension where we restrict our shattered set to be a subspace. Before that, we define subspace and related definitions.

Definition 2.1.4 (Subspace). A subset $U \subseteq \mathbb{F}_2^n$[1] is a subspace if and only if U satisfies the following three conditions.

- $0^n \in U$
- $v, w \in U \Rightarrow v + w \in U$

[1] $\mathbb{F}_2$ denotes the field $\{0,1\}$, and addition, multiplication is defined over modulo 2.

- $\lambda \in \{0, 1\},\ u \in U \Rightarrow \lambda u \in U$

Definition 2.1.5 (Basis). Let $\{e_1, e_2, \dots e_d\}$ be a set of vectors from $\{0, 1\}^n$. They are said to be basis vectors of a subspace U if and only if all e_i are linearly independent, and $\forall v \in U$ can be written as $v = \sum_{i=1}^{d} \lambda_i e_i$ where $\lambda_i \in \mathbb{F}_2$, and $1 \leq i \leq d$.

There can be many sets of basis vectors, and any two sets of basis vectors have the same size. The size of a set of basis vectors is known as *dimension* of the subspace U.

Definition 2.1.6 (Subspace VC-dimension). Fix $U = \{0, 1\}^n$ to be the universe and $\mathcal{F}$ is set of subsets of U. For a subspace $S \subseteq \{0, 1\}^n$, we say S is *subspace-shattered* by family $\mathcal{F}$ if for all subspaces S' of S, $\exists f \in \mathcal{F}$ such that $f \cap S = S'$. The dimension of the largest possible shattered subspace is called the *subspace* VC-*dimension* of the family $\mathcal{F}$.

We now define a different notion of shattering called strong shattering.

Definition 2.1.7 (Strongly Shattered sets)**.** A set S is said to be strongly shattered by the family $\mathcal{F}$ if the subsets of S are appearing in $\mathcal{F}$ with a common set of elements B. That is, there is a $B \subseteq U \setminus S$, such that $\forall S' \subseteq S$, $B \cup S' \in \mathcal{F}$. The largest strongly shattered set is denoted by $\mathsf{VC}_s(\mathcal{F})$.

We immediately observe that $\mathsf{VC}(\mathcal{F}) \geq \mathsf{VC}_s(\mathcal{F})$.

Definition 2.1.8 (Shifting/Compressing a Family $\mathcal{F}$). Let $\mathcal{F}$ be a family of sets over a ground set U. We say that a set $F \in \mathcal{F}$ is shifted by an element $x \in U$ to obtain another set F_x, where F_x is obtained as follows.

$$F_x = \begin{cases} F & , if\ F \setminus \{x\} \in \mathcal{F} \\ F \setminus \{x\} & , otherwise \end{cases}$$

We denote the shifted family as $\mathcal{F}_x$,

$$\mathcal{F}_x = \{F_x \mid F \in \mathcal{F}\} .$$

Notice that, by definition $|\mathcal{F}| = |\mathcal{F}_x|$. In fact, $\mathcal{F}_x$ and $\mathcal{F}$ shatter the same set. We show this in Proposition 3.3.9.

PAC Learning Probably Approximately Correct (PAC) model of learning is a learning model, introduced by Valiant (1984), which is characterized by learning from examples. Consider a universe U and a probability distribution $\mathcal{D}$ defined on U. A concept is a subset $C \subseteq U$. All the elements in C is said to have positive classification and all elements in $U \setminus C$ is said to have negative classification. Equivalently, a classification can be treated as a function $c(x)$ that is 1 if $x \in C$ and 0 otherwise. A concept class $\mathcal{C}$ is set of all possible classifications defined by the problem. Equivalently , $\mathcal{C} \subseteq 2^U$. The learning algorithm calls a function $ORACLE$ that produces a pair $(x, c(x))$, where x is chosen according to the distribution $\mathcal{D}$ and $c(x) = 1$ if $x \in C$ and 0 otherwise. We denote $ORACLE(C, \mathcal{D})$ to specify the concept and distribution under consideration. The concept class $\mathcal{C}$ is said to be PAC learnable if there is an algorithm $\mathcal{A}$, with access to a function $ORACLE(C, \mathcal{D})$, that satisfies the following properties:

1. For every concept $C \in \mathcal{C}$, every distribution $\mathcal{D}$ on U, and every $0 < \epsilon, \delta \leq 1/2$, the number of calls made by the algorithm $\mathcal{A}$ to the function $ORACLE(C, \mathcal{D})$ is polynomial in $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$.
2. The algorithm outputs a hypothesis h with probability at least $1 - \delta$ such that $\Pr_{\mathcal{D}}[h(x) \neq c(x)] \leq \epsilon$.

2.2 Boolean Functions and their Complexity Measures

A set $U = \{0,1\}^n$ is a set of all possible n-bit strings of $\{0,1\}$. A Boolean function $f(x_1, x_2, \ldots, x_n)$ of n variables is a mapping $f : \{0,1\}^n \to \{0,1\}$. An n-dimensional Boolean Hypercube denoted as $\mathcal{B}_n$ is a graph on 2^n vertices, each labeled with an n-bit Boolean string. Two vertices are connected, if and only if they differ exactly at one position. For $x, y \in \{0,1\}^n$ we say $x \prec y$ if $\forall i \in [n], x_i \leq y_i$ where x_i represents i^{th} position bit value of x. Let $x^0, x^1, \ldots x^n \in \{0,1\}^n$ be $n+1$ distinct elements such that $0^n = x^0 \prec x^1 \prec \ldots \prec x^n = 1^n$. Then $x^0, x^1, \ldots x^n \in \{0,1\}^n$ is said to form a *chain* on the Boolean hypercube $\mathcal{B}_n$. Suppose $x, y \in \{0,1\}^n$. Then x and y are said to be related if $x \prec y$ or $y \prec x$. If neither holds then x and y are said to be incomparable elements. A set $S \subseteq \{0,1\}^n$ is called an *antichain* if all the elements in S are incomparable to each other. For example, $S = \{100, 001, 010\}$ is an *antichain* of size 3.

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is said to be *monotone* if and only if $\forall x, y \in \{0,1\}^n, x \prec y \Rightarrow f(x) \leq f(y)$. We now formally define *alternation* of a

Boolean function , which is a measure of non-monotonicity of the Boolean function.

Definition 2.2.1 (Alternation). Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Consider the chain $\mathcal{C}$, $0^n = x^0 \prec x^1 \prec \ldots \prec x^n = 1^n$ on the Boolean hypercube. Alternation on chain $\mathcal{C}$ for function f is denoted as $\mathsf{alt}(f,\mathcal{C})$.

$$\mathsf{alt}(f,\mathcal{C}) = |\{i \mid f(x^{i-1}) \neq f(x^i), x^i \in \mathcal{C}, i \in [n]\}|$$

Alternation of function is denoted as $\mathsf{alt}(f)$ and is equal to maximum alternation over all the chains in the Boolean hypercube.

$$\mathsf{alt}(f) = \max_{\mathcal{C}}\{\mathsf{alt}(f,\mathcal{C})\}$$

Observe that *alternation* of non-constant monotone Boolean functions is indeed 1. An alt-1 family is a family of Boolean functions whose alternation is at most 1. This family has *monotone* functions and negation of *monotone* functions.

For any Boolean function $f : \{0,1\}^n \to \{0,1\}$, there exists a unique n-variate multilinear polynomial $P(x_1, x_2, \ldots, x_n)$ over $\mathbb{R}[x_1, x_2, \ldots, x_n]$ such that $\forall x \in \{0,1\}^n, P(x) = f(x)$.

Definition 2.2.2 (Degree). The degree of a Boolean function f, denoted as $\mathsf{deg}(f)$ is the degree of unique multilinear polynomial $P(x_1, x_2, \ldots, x_n) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ such that $\forall x \in \{0,1\}^n, P(x) = f(x)$.

If the polynomial is over $\mathbb{Z}_k[x_1, x_2, \ldots, x_n]$ for an integer $k > 1$, then we obtain $\mathbb{Z}_k$-degree of f denoted as $\mathsf{deg}_k(f)$.

Definition 2.2.3 (Certificate Complexity). A certificate c of f on input x is defined as a partial assignment $c : P \to \{0,1\}$, where $P \subseteq [n]$ such that f is constant on the restriction c. Equivalently $\forall y \in \{0,1\}^n,\ y|_P = x|_P$[2] $\implies f(y) = f(x)$. If f is always 0 on this restriction, the certificate c is a 0-certificate. If f is always 1, the certificate c is a 1-certificate. Certificate complexity of f on input x denoted by $C(f,x)$ is the minimum $|P|$, such that c is a $f(x)$-certificate.

[2] $y|_P \in \{0,1\}^{|P|}$ such that $\forall i \in P, c(i) = y_i$.

The certificate complexity of f denoted by $C(f)$ is defined as

$$C(f) = \max_{x \in \{0,1\}^n} C(f, x)$$

.

Definition 2.2.4 (Sensitivity). The sensitivity of a function f on an input $x \in \{0,1\}^n$ is defined as $s(f, x) = |\{i \mid f(x) \neq f(x^{\oplus i})\}|$. Sensitivity of function f denoted by $s(f)$ is defined as

$$s(f) \stackrel{\text{def}}{=} \max_{x \in \{0,1\}^n} s(f, x)$$

.

Definition 2.2.5 (Block Sensitivity). The block sensitivity of a function f on an input x denoted by $bs(f, x)$ is defined as the maximum number of disjoint blocks $\{B \mid B \subseteq [n]\}$ such that $f(x) \neq f(x^{\oplus B})$ where $x^{\oplus B}$ denotes all the indices of x in B flipped. Block Sensitivity of the function f denoted by $bs(f)$ is defined as

$$bs(f) \stackrel{\text{def}}{=} \max_{x \in \{0,1\}^n} bs(f, x)$$

Notice that $s(f) \leq bs(f)$.

The *average sensitivity* of a Boolean function is expected sensitivity of f on an input x i.e. $\mathbb{E}_{x \in \{0,1\}^n}[s(f, x)]$. This is also known as *influence* of a function denoted by $I(f)$. For a Boolean function f, $Inf_i(f) = \Pr_{x \in \{0,1\}^n}[f(x) \neq f(x^{\oplus i})]$ i.e. fraction of inputs where i is *influential*. *Influence* or *average sensitivity* can equivalently be written as $I(f) = \sum_{i=1}^{n} Inf_i(f)$.

We will see now how to combine several families together to obtain a new family.

k-closure of Boolean families Given k classes of n-bit Boolean functions $\mathcal{F}_1, \mathcal{F}_2 \ldots, \mathcal{F}_k$, and a fixed Boolean function $f : \{0,1\}^k \to \{0,1\}$. The family of Boolean functions as k-closure is defined as,

$$\mathcal{F}(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k) = \{f(f_1(), ..., f_k()) \mid f_i \in \mathcal{F}_i, i \in [k]\}$$

where same inputs are being fed to all the f_i's. Notice the functions in $\mathcal{F}$ are still functions from $\{0,1\}^n$ to $\{0,1\}$.

Definition 2.2.6 (k-fold union). Let $\mathcal{F} \subseteq 2^{\{0,1\}^n}$. k-fold union family of $\mathcal{F}$ is denoted as $\mathcal{F}^{k\cup}$ and defined as $\mathcal{F}^{k\cup} = \left\{\bigcup_{i=1}^{k} F_i \mid F_i \in \mathcal{F}\right\}$.

Special Boolean Functions We define some specific Boolean function which has appeared in the thesis.

PARITY : We denote PARITY by $\oplus_n$.

$$\textsc{Parity}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & , \ (\sum_{i=1}^{n} x_i) \ mod \ 2 = 1 \\ 0 & , \ otherwise \end{cases}$$

MAJORITY :

$$\textsc{Maj}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & , \ \text{if} \sum_{i=1}^{n} x_i \geq \frac{n}{2} \\ 0 & , \ \text{otherwise} \end{cases}$$

That is, the function evaluates to 1 if and only if the weight of the input is at least $\frac{n}{2}$.

THRESHOLD :

$$\textsc{Th}_n^k(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & , \ \text{if} \sum_{i=1}^{n} x_i \geq k \\ 0 & , \ \text{otherwise} \end{cases}$$

The function evaluates to 1 if the weight of the input is at least k.

k-slice : A function $f : \{0,1\}^n \to \{0,1\}$ is said to be k-slice function if

$$f(x) = \begin{cases} 0 \, , & \sum_{i=1}^{n} x_i < k \\ 1 \, , & \sum_{i=1}^{n} x_i > k \end{cases}$$

SYMMETRIC: A function $f : \{0,1\}^n \to \{0,1\}$ is symmetric if the function is invariant under the permutation of input indices i.e. $f(x) = f(\pi(x))$ where $\pi(x) : \{0,1\}^n \to \{0,1\}^n$ is a bijective function.

READ-ONCE: A Boolean function is said to be *read-once* if there is a Formula (a Boolean circuit where every gate has fan-out at most 1) computing the function f such that every variable appears only once in the formula. Monotone read-once functions have no negations in the formula computing the function.

2.3 Complexity Classes

In this section, we briefly define some complexity classes that appeared in the thesis. Complexity classes are a set of related problems defined in terms of computational difficulty in solving the problem with a computational resource like *time* or *memory* or in terms of *no. of gates* in a circuit. We describe few classes below.

Class **P** is the set of problems that can be computed by a deterministic Turing machine in polynomial time. Class **NP** is the set of problems that can be computed by a non-deterministic Turing machine in polynomial time. A problem X is *hard* for a complexity class C if all the problems in the class C many-one [3] *reduces* to X . The set of problems that are hard for **NP** are **NP**-hard problems. A problem is said to be in class **coNP** if and only if the complement of the problem is in class **NP**. Polynomial hierarchy is a hierarchy of complexity classes that generalizes the class **NP** and **coNP**. $\mathbf{\Sigma_3^p}$ is one such complexity class in the polynomial hierarchy. A language $L \subseteq \Sigma^*$ is said to be in Σ_3^p, if there exists a deterministic Turing machine M running in time polynomial in its first input, such that $x \in L \iff \exists w_1 \forall w_2 \exists w_3 M(x, w_1, w_2, w_3) = 1$ and $|w_1|, |w_2|, |w_3|$ are polynomial in $|x|$. Analogous to class of decision problems, we define complexity classes for function computation problems. A natural extension of **P** is **FP** which is defined as $\mathsf{FP} = \{f \mid f : \Sigma^* \to \mathbb{N}, \text{for any } x \in \Sigma^*, f(x) \text{ can be written in } poly(|x|) \text{ time}\}$. In a similar way we can define $\mathsf{FP}^{\Sigma_3^p}$ which will be analogous to complexity class Σ_3^p.

Definition 2.3.1 (LOGNP Class)**.** The class of decision problems expressible in logical form as, "the set of I for which there exists a subset $S = \{s_1, s_2, \ldots, s_{\log n}\} \subseteq [n]$, such that $\forall x\ \exists y$ such that $\forall j \in S$, the predicate $\phi(I, s_j, x, y, j)$ holds where x and y are logarithmic-length strings, and ϕ is computable in P".

We now see a complexity class $\mathcal{AM}$ based on Arthur-Merlin protocol. Arthur tosses some random coins and sends the outcome of all his coin tosses to Merlin.

[3] Suppose A and B are two formal languages over the alphabets Σ and Γ respectively. We say A many-one reduces to B if there exists a total computable function $f : \Sigma^* \to \Gamma^*$, such that for any string $w \in A \Leftrightarrow f(w) \in B$.

Merlin responds with a purported proof, and Arthur deterministically verifies the proof. Formally,

Definition 2.3.2 ($\mathcal{AM}$ Class)**.** A language $L \in \mathcal{AM}$ if there exists a deterministic algorithm $\mathbb{V}$ running in polynomial time (in the length of its first input) such that:

- If $x \in L$, then $\forall r\ \exists \pi$ such that $\mathbb{V}(x, \pi, r) = 1$
- If $x \notin L$, then $\forall \pi$ we have $\mathsf{Pr}_r[\mathbb{V}(x, \pi, r) = 1] \leq \frac{1}{2}$

CC^0 is the class of Boolean function families which can be computed by circuit families of polynomial size, constant depth which uses MOD_q gates for any integer q (which is independent of n). The class AC^0 is the class of Boolean function families which can be computed by constant depth polynomial size circuit families using the gates $\wedge$, $\vee$, and $\neg$ gates.

We refer the readers to the textbook by Arora and Barak (2009) and Kozen (2006) for more details.

CHAPTER 3

Literature Survey

In this chapter, we give a brief overview of Boolean functions complexity and known results and connection of Learnability of Boolean functions and VC-dimension . Firstly, we look at a measure of non-monotonicity called *alternation* (see Definition 2.2.1) and then survey known relations between various measures of complexity of a Boolean function. Then, we see the connection of VC-dimension in learnability of Boolean family and state the lower and upper bound in Section 3.2. Then, we see various properties of VC-dimension of a family and see some results relevant in the study of the thesis. We then look at a family of k-union (Section 3.3.2) and show the literature and known bounds for it. We also look at a special family called *s-extremal* family (Section 3.4.2). Finally, in Section 3.5, we look at the complexity of computing VC-dimension in different variants and conclude with some new applications of VC-dimension .

3.1 Boolean Function Complexity

A Boolean function $f(x_1, x_2, \ldots, x_n)$ of n variables is a mapping $f : \{0,1\}^n \to \{0,1\}$. We say $f(x)$ evaluates to *True* if the value $f(x) = 1$ and *False* if $f(x) = 0$. All the inputs or vectors for which $f(x) = 1$ are called positive inputs (and negative inputs when $f(x) = 0$). The name Boolean function comes from Boolean logic invented by *George Bool*. Since total number of inputs are 2^n, there are 2^{2^n} Boolean functions. Some examples of Boolean functions are *AND* ($f(x) = 1 \Leftrightarrow$ if all bits are 1), *Parity* ($f(x) = 1$ if no. of 1's is odd in x), *Majority* ($f(x) = 1$ if no. of 1's is at least $n/2$) etc. For more details we refer the readers to the textbook by Crama and Hammer (2011).

There are many algebraic and combinatorial Boolean function complexity measures such as *degree, sensitivity, block sensitivity, influence, alternation* etc. which we have defined in Chapter 2.

3.1.1 Measures of Non-monotonicity

We look now at various measures of non-monotonicity to understand non-monotonicity of a Boolean function.

Orientation A function $f : \{0,1\}^n \to \{0,1\}$ is said to have orientation $\beta \in \{0,1\}^n$ if there exists a monotone function $h : \{0,1\}^{2n} \to \{0,1\}$ such that $\forall x \in \{0,1\}^n,\ f(x) = h(x, x \oplus \beta)$. Observe that for a monotone function 0^n is an orientation. The weight of β ($wt(\beta)$) indicates how close a function is to being monotone. This was studied by Koroth and Sarma (2017) and we refer the same to interested readers.

An implicant of a Boolean function $f(x_1, x_2, \ldots, x_n)$ is a nonzero term p such that $p(a) \leq f(a)$ for all $a \in \{0,1\}^n$. An implicant p of a Boolean function f is a prime implicant, if upon removing all the occurrences of any particular literal from p, the resulting term is not an implicant of f anymore.

Negation Width A *DeMorgan*[1] circuit computing a monotone Boolean function f has negation width w if for every prime implicant p of f, the circuit produces either p or some extension containing at most w negated variables. Extension of a prime implicant p is a nonzero term of the form $p.r$ where $r = \overline{x_{i_1}}.\overline{x_{i_2}}.\ldots.\overline{x_{i_t}}$ consists of only negated literals. A monotone Boolean circuit has negation width 0. Jukna and Lingas (2019) gave a relation between a circuit with negation width w and size of monotone circuit computing the function.

Monotone Decision Trees Another measure of non-monotonicity is *monotone decision trees* which are decision trees where each query is a monotone function on the input bits. Monotone decision tree height is shown to be related with $\mathsf{alt}(f)$ by Amireddy *et al.* (2020) and we refer the readers the same for more information.

Our interest of study is one particular measure of non-monotonicity called *alternation*. A function $f : \{0,1\}^n \to \{0,1\}$ is said to be *monotone* if and only if $\forall x, y \in \{0,1\}^n, x \prec y \Rightarrow f(x) \leq f(y)$. Notice that *alternation* of a monotone function is 1. *Alternation* of a Boolean function can be at most n. Higher the *alternation* of a Boolean function, higher the *non-monotonicity*. A close variant of alternation called *decrease* (denoted by $\mathsf{dec}(f)$, where only the indices $i \in [n]$ such that $f(x_{i-1}) > f(x_i)$ are counted) was originally introduced by Markov (1958) while more recent literature Blais

[1] A circuit with all negation at the leaves

et al. (2015); Lin and Zhang (2017); Dinesh and Sarma (2018) uses alternation as the measure of non-monotonicity. Note that, for all Boolean functions f, $\mathsf{d}ec(f) = \lfloor \frac{\mathsf{alt}(f)}{2} \rfloor$ and functions having decrease as zero are exactly monotone functions. We look at the following characterization of *non-monotone* functions due to Blais *et al.* (2015).

Lemma 3.1.1 (Blais *et al.* (2015))**.** *Let $f : \{0,1\}^n \to \{0,1\}$. Then there exists $k = \mathsf{alt}(f)$ monotone functions $g_1, \ldots, g_k$ each from $\{0,1\}^n$ to $\{0,1\}$ such that*

$$f(x) = \begin{cases} \oplus_{i=1}^{k} g_i & , \textit{if } f(0^n) = 0 \\ \neg \oplus_{i=1}^{k} g_i & , \textit{if } f(0^n) = 1 \end{cases}$$

We give a brief sketch of the proof of the above lemma. The decomposed monotone functions g_i's are defined as $g_i = \{x \in \{0,1\}^n \mid alt(f,x) \leq i-1\}$ where $\mathsf{alt}(f,x)$ denotes maximum alternation for any chain in f starting at x. From definition observe that $g_{i+1} \Rightarrow g_i$ (in set theory terms $g_{i+1} \subseteq g_i$). These decomposed monotone functions or boundary functions divides the entire Boolean Hypercube $\mathcal{B}_n$ of f in k-boundaries and their parity is exactly the value of the function at all points or inputs.

3.1.2 Relation between Complexity Measures

We defined several complexity measures of a Boolean function in Chapter 2. It turns out that all these measures are polynomially related and we summarize them below. In Chapter 5 we show a connection between VC-dimension and *Certificate complexity* ($C(f)$) and also show a characterization of a highly sensitive function. Relation between various Boolean function complexity measures is summarized below in Table 3.1. The relations hold for any Boolean function.

Relation	Citation
$s(f) \leq bs(f) \leq C(f)$ $bs(f) \leq D(f)$	Folklore
$D(f) \leq s(f)bs(f)^2$ $C(f) \leq s(f)bs(f)$	Nisan (1991)
$bs(f) \leq 2\mathsf{deg}(f)^2$	Nisan and Szegedy (1992)
$D(f) \leq \mathsf{deg}(f)^3$ $\mathsf{deg}(f) \leq D(f) \leq C(f)^2$	Midrijanis (2004)
$bs(f) \leq \mathsf{deg}(f)^2$	Tal (2013)
$bs(f) = O(\mathsf{alt}(f)^2 s(f))$	Lin and Zhang (2017)
$bs(f) \leq s(f)^4$	Huang (2019)

Table 3.1: Relation between the complexity measures considered

3.2 Learnability of Boolean Family

In this section, we look at the learnability of family of Boolean functions in particular. There are many classes of Boolean functions and they abstract out the data obtained from the real world. The aim has always been to estimate the properties of an entire sample space using as few data or samples as possible. We look at sample complexity of Bounds in Valiant (1984) PAC model of learnability.

The Learning Model We saw the definition of PAC Learnability in Chapter 2. The learning algorithm requires a training set or examples of the form $(x_1, c(x_1))$, $(x_2, c(x_2)) \ldots, (x_m, c(x_m))$ where x_i's are chosen according to some distribution $\mathcal{D}$. c is the target concept and $c(x_i) = 1$ if and only if $x_i \in c$. The algorithm also has a collection of hypothesis called the Hypothesis Class or Concept Class $\mathcal{C}$. The algorithm outputs a hypothesis h and its correctness is evaluated with respect to closeness with the target concept. The aim is to minimize the sample complexity with a better approximation of the hypothesis produced and the target concept. Now we will state the exact sample complexity bounds due to Ehrenfeucht *et al.* (1989) and Blumer *et al.* (1989) that one gets from VC-dimension of a family.

Theorem 3.2.1 (Ehrenfeucht *et al.* (1989)). *Consider a family $\mathcal{F}$ such that* $\mathsf{VC}(\mathcal{F}) \geq 2$

and $0 < \epsilon \leq \frac{1}{8}, 0 < \delta \leq \frac{1}{100}$. *Then any* (ϵ, δ) *algorithm* $\mathcal{A}$ *for* $\mathcal{F}$ *must use sample size* $\Omega(\frac{\mathsf{VC}(\mathcal{F})}{\epsilon})$.

We give a brief sketch of the proof of the above Theorem. We will refer to some of these in later discussions. Let X be the universe. Consider a set $X_0 = \{x_0, x_1, \ldots, x_d\} \subseteq X$ shattered by $\mathcal{F}$, where $\mathsf{VC}(\mathcal{F}) = d + 1$. Define the following distribution P on X.

$$P(x) = \begin{cases} 1 - 8\epsilon, & x = x_0 \\ \frac{8\epsilon}{d}, & x = x_i,\ 1 \leq i \leq d \\ 0, & otherwise \end{cases}$$

Define $\mathcal{F}_0 = \{\{x_0\} \cup T \mid T \subseteq \{x_1, x_2, \ldots, x_d\}\}$. Since P is 0 except on X_0, we may assume that $X = X_0$ and $\mathcal{F} = 2^{X_0}$. Hence, $\mathcal{F}_0 \subseteq \mathcal{F}$. Let m be the number of samples used by the algorithm. Define a set $S \subset X^m$,

$$S = \{\vec{x} \in X^m \mid \vec{x} \text{ has at most } d/2 \text{ distinct elements from } \{x_1, x_2, \ldots, x_d\}\}$$

The following two lemmas establishe the desired bound.

Lemma 3.2.2. *Let* U *be the uniform distribution on* $\{0, 1\}$. *Let* h *be the produced hypothesis by the learning algorithm. Then there exists a* $c_0 \in \mathcal{F}_0$ *satisfying*

$$\mathsf{Pr}_{\vec{x} \in P^m\ r \in U^k}[h \triangle c_0 > \epsilon] > \frac{1}{7} P^m(S)$$

where $h \triangle c_0$ *is fraction of inputs where* h *and* c_0 *disagree.*

Lemma 3.2.3. *If* $\delta \leq \frac{1}{100}$. *If* $m = \frac{d}{32\epsilon}$ *then* $P^m(S) > 7\delta$.

Theorem 3.2.4 (Blumer *et al.* (1989))**.** *Let* $\mathcal{F}$ *be a family of subsets of universe* U *such that* $\mathsf{VC}(\mathcal{F}) = d$. *Then there exists an* (ϵ, δ)*-learning algorithm* $\mathcal{A}$ *for a family* $\mathcal{F}$ *that uses at most* $\mathcal{O}(\frac{d}{\epsilon} \ln \frac{1}{\epsilon})$.

Thus, computing VC-dimension of families immediately answers whether there exists an efficient learning algorithm for $\mathcal{F}$ or family is efficiently learnable with provable optimal sample complexity.

Dinesh and Sarma (2018) gives a bound on sample complexity of learning algorithm for families of Boolean functions parameterized by *alternation*.

Theorem 3.2.5 (Dinesh and Sarma (2018))**.** *Let*

$$\mathcal{A} = \{f : \{0,1\}^n \to \{0,1\} \mid alt(f) \leq k\}$$

where $alt(f)$ is the alternation of the function f. We have an ϵ error learning algorithm from random examples with a runtime of $\mathcal{O}(n^{alt(f)\times \mathsf{deg}_2(\mathsf{f})^2/\epsilon})$.

A bound on running time also implies an upper bound on the sample complexity as number of queries is limited by running time of the algorithm. We remark here that the underlying distribution of the learning algorithm is the uniform distribution.

3.3 Known Bounds and Properties of VC-dimension

In this section, we revisit the definition of VC-dimension and see some more properties of VC-dimension of a family.

Let U be a universe and $\mathcal{F}$ be a set of subsets of U. The family $\mathcal{F}$ is said to shatter a subset $S \subseteq U$ if for all $S' \subseteq S$, there is a $F \in \mathcal{F}$ such that $S \cap F = S'$. The VC-dimension of $\mathcal{F}$ is said to be the largest d such that there is a set S of size d shattered by the family. We will see some results that characterize the VC-dimension of a family with respect to the size of the class. These results are well known but we present our own proof for completeness.

Observation 3.3.1. *Let $\mathcal{F}$ be family which shatters a set S. Then the following holds from the definition of* VC*-dimension .*

1. $\mathsf{VC}(\mathcal{F}) \geq |S|$
2. *$\mathcal{F}$ also shatters all subsets of S.*
3. $\mathcal{F} \geq 2^{|S|}$
4. $\mathcal{F} \subseteq \mathcal{G} \Rightarrow \mathsf{VC}(\mathcal{F}) \leq \mathsf{VC}(\mathcal{G})$

Proposition 3.3.2. *For any family of Boolean functions $\mathcal{F}$,* $\mathsf{VC}(\mathcal{F}) \leq \log |\mathcal{F}|$.

Proof. The proof follows from the fact that if the $\mathsf{VC}(\mathcal{F}) = k$, then there is a set $S \subseteq U$ of size k which is shattered, and hence there must be at least one $F \in \mathcal{F}$ for each subset of S. Thus $|\mathcal{F}| \geq 2^k$ which implies the bound. □

There is a similar kind of result for the lower bound of VC-dimension for a family. Before seeing that we first see *projection* of a set S on a family $\mathcal{F}$ denoted by $\Pi_F(S)$. This is defined as

$$\Pi_{\mathcal{F}}(S) = \{F \cap S \mid F \in \mathcal{F}\}.$$

This is also sometime termed as *growth function*. Observe that if a set S is shattered by the family $\mathcal{F}$ then $\Pi_{\mathcal{F}}(S) = 2^S$. For an integer $m > 0$, define $\Pi_{\mathcal{F}}(m) = \max_{S \subseteq U, |S|=m} |\Pi_{\mathcal{F}}(S)|$. Notice that if $\mathsf{VC}(\mathcal{F}) = d$ then $\Pi_{\mathcal{F}}(d) = 2^d$, and in fact, $\Pi_{\mathcal{F}}(m) = 2^m$ for all $m \leq d$. In other words, $\mathsf{VC}(\mathcal{F})$ is the maximum value of m such that $\Pi_{\mathcal{F}}(m) = 2^m$. Now we state a lemma about value of $\Pi_{\mathcal{F}}(m)$ for $m > d$. This Lemma is popularly known as *Sauer-Shelah* lemma and was first proved by Vapnik and Chervonenkis (1971) in their original paper on VC-dimension and also independently by Sauer (1972) and Shelah (1972) among other authors.

Lemma 3.3.3 (Sauer-Shelah Lemma). *Let $\mathcal{F}$ be a family and $\mathsf{VC}(\mathcal{F}) \leq d$, where d is at most the size of the universe U and $(d \geq 3)$,*

$$\Pi_{\mathcal{F}}(m) \leq 1 + \binom{m}{1} + \binom{m}{2} + \ldots + \binom{m}{d} \leq \left(\frac{em}{d}\right)^d \leq m^d$$

We now show the lower bound on VC-dimension of a family.

Proposition 3.3.4. *Let $\mathcal{F}$ be a family such that $\mathsf{VC}(\mathcal{F}) \geq 3$. Then $\mathsf{VC}(\mathcal{F}) \geq \frac{\log|\mathcal{F}|}{\log|U|}$.*

Proof. At $m = |U|$ all the elements in $\mathcal{F}$ must give different vectors in the projection, and hence $\Pi_{\mathcal{F}}(m) = |\mathcal{F}|$. Suppose $\mathsf{VC}(\mathcal{F}) \leq d$, then

$$|\mathcal{F}| = \Pi_{\mathcal{F}}(m) \leq m^d \leq |U|^d$$

Thus,

$$d \geq \frac{\log|\mathcal{F}|}{\log|U|}$$

□

We have the following corollary on the lower bound of VC-dimension of a family of Boolean functions where $U = \{0,1\}^n$.

Corollary 3.3.5. *Let $\mathcal{F}$ be a family of Boolean functions. Then $\mathsf{VC}(\mathcal{F}) \geq \frac{\log|\mathcal{F}|}{n}$.*

We collect (in Table 3.2) known VC-dimension bounds of various Boolean function classes. All functions are of the form $f : \{0,1\}^n \to \{0,1\}$ except $\mathcal{C}_{n,k}$ which has families of Boolean functions $g : \{-1,1\}^n \to \{-1,1\}$.

Function Family	**UB**	**LB**	**Remarks**
Monotone Functions	$\binom{n}{n/2}$	$\binom{n}{n/2}$	Procaccia and Rosenschein (2006)
Monomials	n	n	Natschläger and Schmitt (1996)
Monotone Monomials	n	n	Natschläger and Schmitt (1996)
DNFs with terms size k	$O(n^k)$	$\Omega(n^k)$	Ehrenfeucht *et al.* (1989)
$\mathcal{C}_{n,k}$	$2nk$	$\max\{n,k\}$	Mixon and Peterson (2015)
k-Decision Lists	$O(n^k)$	$\Omega(n^k)$	Rivest (1987)
Symmetric Functions	$n+1$	$n+1$	Ehrenfeucht *et al.* (1989)

Table 3.2: Bounds on VC-dimension of Families of Functions

In Table 3.2, $C_{n,k}$ is a family of Boolean functions defined over $\{-1,1\}^n$ domain instead of $\{0,1\}^n$. A function $g : \{-1,1\}^n \to \{-1,1\}$ is said to be in $\mathcal{C}_{n,k}$, if there are subsets $S_1, S_2, \ldots S_k \subseteq [n]$ and coefficients $a_1, a_2, \ldots a_k \in \mathbb{R}$ such that :

$$g(x) = \mathsf{sign}\left(\sum_{i=1}^{n} a_i \prod_{j\in S_i} x_i\right)$$

3.3.1 Properties of VC Dimension

We collect here some interesting properties of VC-dimension of a family.

Lemma 3.3.6. *Let $\mathcal{F}$ be a family such that its* $\mathsf{VC}(\mathcal{F}) = d$. *Consider a set $H \subseteq \mathcal{U}$ and the family $\mathcal{F}_H = \{F \triangle H \mid F \in \mathcal{F}\}$. Then* $\mathsf{VC}(\mathcal{F}) = \mathsf{VC}(\mathcal{F}_H)$.

We did not find a proof of the above lemma in our literature survey. Hence we provide a proof of our own for completeness.

Proof. Lower Bound: Let S be the largest set shattered by $\mathcal{F}$. We will show that $\mathcal{F}_H$ also shatters the set S. These are the following cases which arise while shattering. The notation F_T implies the concept $F_T \in \mathcal{F}$ such that $F_T \cap S = T$.

1. Consider that $S' \subseteq S$ and $S' \cap H = \phi$. In this case $F' = F_{(H\cap S)\cup S'}$.
2. For the case when $S' \subseteq S$ and $S' \subseteq H$, define $W = (H \cap S) \setminus S'$. In this case, $F' = F_W$

3. For the case $S' \subseteq S$ and $S' \cap H \neq \phi$, suppose $S' = X \cup Y$ such that $X \cap H = \phi$ and $Y \subseteq H$. Define $W = (H \cap S) \setminus S'$. In this case, $F' = F_{W \cup X}$.

For each of the above case, $(F' \triangle H) \cap S = S'$. Thus we can obtain all the subsets of S and hence $\mathcal{F}_H$ shatters S. We give below a Venn-diagram (Figure 3.1) to understand the construction above.

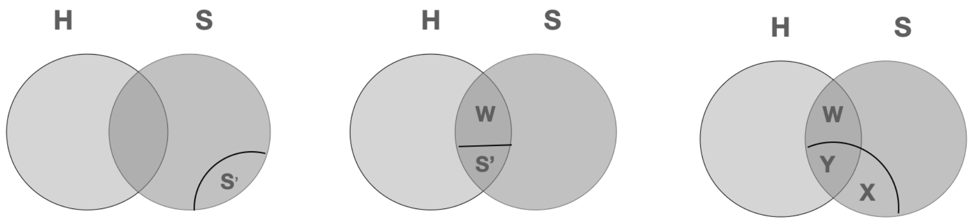

Figure 3.1: Venn Diagram of H and S

Upper Bound: We show this by contradiction. Suppose that, $\mathcal{F}_H$ shattered a set of size greater than d. Then using the method above, we will also be able to show that $\mathcal{F}$ shattered the same set. This is due to bijective nature of the construction. Observe that $(\mathcal{F}_H)_H = \mathcal{F}$ by the definition of symmetric difference. Thus, if $\mathcal{F}$ shatters a set then we know how the same set will be shattered by $\mathcal{F}_H$ and vice versa.

□

We have the following observation from the above lemma. Let $\mathcal{F}$ be a family and U be the universe. We define $\overline{\mathcal{F}}$ as follows.

$$\overline{\mathcal{F}} = \{U \setminus F \mid F \in \mathcal{F}\}$$

Observation 3.3.7 (Negation of a family)**.** $\mathsf{VC}(\mathcal{F}) = \mathsf{VC}(\overline{\mathcal{F}})$

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two families.

Observation 3.3.8 (DeMorgan's Law)**.** $\mathsf{VC}(\mathcal{F}_1 \cup \mathcal{F}_2) = \mathsf{VC}(\overline{\mathcal{F}_1 \cup \mathcal{F}_2}) = \mathsf{VC}(\overline{\mathcal{F}_1} \cap \overline{\mathcal{F}_2})$

We will see now another interesting property. Let $\mathcal{F}$ be a family and $\mathcal{F}_x$ be the *shifted* family (see Definition 2.1.8).

Proposition 3.3.9. *For any set $S \subseteq X$:*

$$\mathcal{F}_x \textit{ shatters } S \Rightarrow \mathcal{F} \textit{ shatters } S$$

Proof. Since $\mathcal{F}_x$ shatters S, we have $\forall S' \subseteq S, \exists F' \in \mathcal{F}_x$ such that $F' \cap S = S'$. Due to the shifting of the family, if $x \in F'$ then both the sets F' and $F' \setminus \{x\}$ are in the family $\mathcal{F}$ and if $x \notin F'$ then either F' or $F' \cup \{x\}$ is in the family $\mathcal{F}$. We will use this information to show that $\mathcal{F}$ shatters the set S.

Consider a subset $S' \subseteq S$ and $x \in S'$. Then $x \in F'$ and hence $\mathcal{F}$ can obtain the subset S'. Similarly, when $x \notin S'$ then x may or may not be in F'. In any case, we can see above that $\mathcal{F}$ can obtain the subset S'. Thus $\mathcal{F}$ shatters the set S.

□

Observation 3.3.10. *The family $\mathcal{F}$ can be shifted multiple times to obtain a new family $\mathcal{F}'$ which is subset closed i.e. if $A \in \mathcal{F}'$ then all subsets of A are also in the family $\mathcal{F}'$. The largest cardinality set is the* VC*-dimension of family $\mathcal{F}$.*

Proof. Suppose there is a set $S = A \cup \{x\}$ such that A is not in the family. In that case, we can shift S by x to make it subset closed. We repeat this process for all such sets to obtain a subset closed family in the end. □

Now we will look at a result due to Ben-David (2015) which characterizes a family with VC-dimension at most 1. If the sets in a family are linearly ordered by inclusion then VC-dimension of the family is 1. For the reverse direction of this statement, the author shows that the family has a very simple description and one does not necessarily need a linear order. Instead, the partial ordering being a *tree* suffices that VC-dimension is at most 1.

Tree Ordering A partial order $\preceq$ over the universe X is a *tree ordering* if $\forall x \in X$, the initial segment ($I_x = \{y \mid y \preceq x\}$) is linearly ordered under $\preceq$.

Theorem 3.3.11 (Ben-David (2015))**.** *Let $\mathcal{H}$ be a family over some universe X. The following statements are equivalent.*

1. $\mathsf{VC}(\mathcal{H}) \leq 1$
2. *There exists a tree ordering over universe X and a $f : X \to \{0, 1\}$ such that every element of $\mathcal{H}_f$*[2] *is an initial segment under that ordering.*

[2] $\mathcal{H}_f = \{h \triangle f \mid h \in \mathcal{H}\}$ where $\triangle$ is symmetric difference of two sets.

3.3.2 VC-dimension of k-fold Union

We defined k-union of families in the Chapter 2 (see Defn 2.2.6). We survey some known bounds of VC-dimension of k-fold union.

Let $\mathcal{F}$ be the family and $\mathcal{F}^{k\cup}$ be the k-union of the family. Suppose VC-dimension of the family $\mathcal{F}$ is d. How large can the VC-dimension of $\mathcal{F}^{k\cup}$ be?

Theorem 3.3.12 (Blumer *et al.* (1989)). *Let $\mathcal{F} \subseteq 2^X$ be a concept class of finite VC-dimension $d \geq 1$. For all $k \geq 1$ let $\mathcal{F}^{k\cup} = \left\{\cup_{i=1}^{k} F_i \mid F_i \in \mathcal{F}\right\}$. Then for all $k >= 1$, the VC-dimension of $\mathcal{F}^{k\cup}$ is less than $2dk\log(3k)$.*

In the same paper, Blumer *et al.* (1989) shows a similar result for a family $\mathcal{F}^{k\cap}$ (union replaced with the intersection as defined above).

Reyzin (2006) gives a construction of a family $\mathcal{F}^{k\cup}$ with VC-dimension at least $\frac{8}{5}dk$ thus proving a lower bound on k-fold union of family. The author also remarks that their method cannot match the upper bound due to Blumer *et al.* (1989) and thus a quest for settling the gap between lower bound and upper bound remains. The gap was settled by *Eisentat and Angluin* in 2007. Eisenstat and Angluin (2007) shows that $\mathcal{O}(dk\log k)$ bound is asymptotically tight for k-union family by constructing a family with VC-dimension at least $\Omega(dk\log k)$.

Theorem 3.3.13 (Eisenstat and Angluin (2007)). *There exists real $\alpha > 0$ such that for all $d \geq 5$ and for all $k \geq 1$, there is a concept class $\mathcal{C}$ of VC-dimension at most d such that $\mathcal{C}^{k\cup}$ has VC-dimension at least $\alpha.dk\log k$.*

We describe the idea used in the construction of the family which achieves the lower bound and a brief sketch of the proof of the theorem. Let n be a positive integer and let $U = \{1, 2, \ldots, n2^n\}$ be the universe. Construct a random concept class $\mathcal{C} = \bigcup_{b=0}^{\lceil\log n\rceil} \mathcal{C}_b$ where $\mathcal{C}_b = \{c_1, c_2, \ldots, c_t\}$ where $t > 0$ and $c_i \subseteq U$ are random a-point [3] concepts. The concept class $\mathcal{C}$ satisfies the following three properties. We leave out the proof and only mention the properties.

1. Each concept in $\mathcal{C}$ appears at most once.
2. For $1 \leq b \leq \lceil\log n\rceil$, and $\forall S \subseteq U$ such that $|S| > n2^{n-b}$, there exists a concept $c \in \mathcal{C}$ such that $c \subseteq S$ and $|c| \geq n/b$.

[3] a-point set: Set of size a.

3. No two concepts in $\mathcal{C}$ intersect in 5 or more points[4].

We need the following 2 lemmas which mention the significance of these properties.

Lemma 3.3.14. *If $\mathcal{C}$ satisfies Properties 1 and 2, then for all $k \geq 3.2^n$, $\mathcal{C}^{k\cup}$ shatters U.*

Lemma 3.3.15. *If concept class $\mathcal{C}$ satisfies property 3, then* $\mathsf{VC}(\mathcal{C}) \leq 5$.

Proof. Suppose, $\mathcal{C}$ shatters a set $S \subseteq U$. Consider $w \in S$. Then there exists two concepts c_1 and c_2 such that $S = c_1 \cap S$ and $S \setminus \{w\} = c_2 \cap S$. Thus $S \setminus \{w\} = c_1 \cap c_2$ and we know from property 3 that no two concept intersects at 5 or more points. Hence, $|S| \leq |c_1 \cap c_2| + 1$, which gives $|S| \leq 5$. □

We need the below Lemma by Reyzin (2006) and Eisenstat and Angluin (2007) for the final argument which increases the VC-dimension of a family and its k-fold union proportionally.

Lemma 3.3.16 (Tagged ℓ-fold union). *If a concept class $\mathcal{C}$ over some universe U has* VC*-dimension d and $\mathcal{C}^{k\cup}$ has* VC*-dimension d', then for all $\ell \geq 1$, there exists a concept class $\mathcal{D}$ with universe $[\ell] \times U$ such that $\mathcal{D}$ has* VC*-dimension ℓd, and $\mathcal{D}^{k\cup}$ has* VC*-dimension $\ell d'$.*

The constructed family $\mathcal{C}$ has VC-dimension at most 5 and its k-fold union has VC-dimension at most $\Theta(k \log k)$. Now we use Lemma 3.3.16 to increase the VC-dimension to our desired value. We obtain a class $\mathcal{D}$ for $\ell = \frac{d}{5}$ with $\mathsf{VC}(\mathcal{D})$ at most 5ℓ and $\mathsf{VC}(\mathcal{D}^{k\cup}) = \frac{d}{5}\Theta(k \log k)$, which proves the bound.

3.4 Sumset Lemma and s-extremal Family

A concept class in this context is a subset $A \subseteq \{0,1\}^n$. Each n-bit vector $x \in A$ can be interpreted as a set $S = \{i \in [n] \mid x_i = 1\}$. Thus, the set A can be equivalently considered as a family of subsets of $[n]$.

[4] points: elements of the set

3.4.1 Interpolation Degree

For any function $f : \{0,1\}^n \rightarrow \{0,1\}$, there is a unique multilinear polynomial $P \in \mathbb{F}_2[x_1, x_2, \ldots, x_n]$ such that $\forall x \in \{0,1\}^n,\ P(x) = f(x)$. But for a partial function $f : A \rightarrow \{0,1\}$ there are $2^{2^n - |A|}$ multilinear polynomials whose restriction to A computes f. *Interpolation-degree* denoted by $\mathsf{int\text{-}deg}$ is defined as the smallest d such that any function $f : A \rightarrow \{0,1\}$ can be computed by a polynomial of degree at most d. Clearly, $\mathsf{int\text{-}deg}$ is an integer between 0 and n. We denote $\mathsf{VC}(A)$ as the VC-dimension of the family corresponding to $A \subseteq \{0,1\}^n$. Then the following connection to VC-dimension is known due to Gurvits (1997).

Lemma 3.4.1 (Gurvits (1997))**.** *Let* $A \subset \mathbb{F}_2^n$. *Then* $\mathsf{int\text{-}deg}(A) \leq \mathsf{VC}(A)$.

We give a brief sketch of the proof of this lemma. Since the set of all multilinear monomials span the set of functions $f : A \rightarrow \{0,1\}$, it suffices to show that any monomial can be represented as a polynomial of degree at most $d = \mathsf{VC}(A)$. This follows from the fact that if a monomial $x_S = \Pi_{i \in S} x_i$ has degree larger than $d = \mathsf{VC}(A)$, then A cannot shatter the set S. Let us revisit another formulation of Sauer-Shelah Lemma 3.3.3 to see a converse statement by Dvir and Moran (2018).

Lemma 3.4.2. *Let A be a family such that* $\mathsf{VC}(A) \leq d$. *Then* $|A| \leq \sum_{i=0}^{d} \binom{n}{i}$.

Consider a family $A \triangle A$ defined as $\{S \triangle T \mid S, T \in A\}$ where $\triangle$ denotes symmetric difference operation. The above Lemma 3.4.2 gives a bound for size of the family $A \triangle A$ as

$$|A \triangle A| \leq \left(\sum_{i=0}^{d} \binom{n}{i}\right) \cdot \left(\sum_{i=0}^{d} \binom{n}{i}\right)$$

Dvir and Moran (2018) shows a converse result for the same i.e. an upper bound on VC-dimension of $A \triangle A$ implies an upper bound on $|A|$.

Theorem 3.4.3 (Theorem 2, Dvir and Moran (2018))**.** *Consider a family A such that* $\mathsf{VC}(A \triangle A) \leq d$. *Then*

$$|A| \leq 2 \sum_{i=0}^{d/2} \binom{n}{\lfloor d/2 \rfloor}$$

.

Dvir and Moran (2018) also show a stronger statement in terms of *interpolation-degree*.

Theorem 3.4.4 (Theorem , Dvir and Moran (2018)). *Let* $d \leq n$ *and* $A \subset \{0,1\}^n$ *satisfying* $|A| > 2\sum_{i=0}^{d/2}\binom{n}{\lfloor d/2 \rfloor}$. *Then* $\texttt{int-deg}(A \triangle A) > d$.

3.4.2 Shattering-Extremal Family

We now see a special class of family called *Shattering-Extremal* family or *s-extremal* family in short. Let $\mathcal{F}$ be family and $Sh(\mathcal{F})$ denote the set of sets shattered by $\mathcal{F}$. The following lemma gives a bound on the size of $Sh(\mathcal{F})$.

Lemma 3.4.5. $|Sh(\mathcal{F})| \geq |\mathcal{F}|$.

This lemma is often referred as *Sauer's* inequality and proved by Pajor (1985), Sauer (1972) and Shelah (1972). The case of equality is of our interest. The family $\mathcal{F}$ is said to be *s-extremal* if and only if $|Sh(\mathcal{F})| = |\mathcal{F}|$.

Union of s-extremal Families Let $\mathcal{F}$ be an *s-extremal* family. Consider the following families constructed from $\mathcal{F}$ as follows.

$$\mathcal{F}_0^i = \{F \mid F \in \mathcal{F}, i \notin F\}$$

$$\mathcal{F}_1^i = \{F \setminus \{i\} \mid F \in \mathcal{F}, i \in F\}$$

Proposition 3.4.6. *If* $\mathcal{F}$ *is* s-extremal *then so is* $\mathcal{F}_1^i$ *and* $\mathcal{F}_0^i$.

Proof. As described above, the family $\mathcal{F}$ is decomposed into two families. We observe that $Sh(\mathcal{F}_1^i) \cup Sh(\mathcal{F}_0^i) \subseteq Sh(\mathcal{F})$ and $|\mathcal{F}_0^i| + |\mathcal{F}_1^i| = |\mathcal{F}|$. Also notice that if $S \in Sh(\mathcal{F}_1^i) \cap Sh(\mathcal{F}_0^i)$ then $S \cup \{i\} \in Sh(\mathcal{F})$. This is because all the subsets of S containing i can be obtained from $\mathcal{F}_1^i$ (sets from the original family $\mathcal{F}$) and the rest from $\mathcal{F}_0^i$. Now we have

$$|Sh(\mathcal{F})| = |\mathcal{F}| = |\mathcal{F}_0^i| + |\mathcal{F}_1^i| \geq |Sh(\mathcal{F}_1^i)| + |Sh(\mathcal{F}_0^i)| \tag{3.1}$$

Suppose that, without loss of generality $\mathcal{F}_0^i$ was not *s-extremal* which means $|Sh(\mathcal{F}_0^i)| > |\mathcal{F}_0^i|$, then it would imply that $|Sh(\mathcal{F}_1^i)| < |\mathcal{F}_1^i|$, contradicting Lemma 3.4.5. □

The reverse of the above proposition need not be true i.e. union of two *s-extremal* families need not be *s-extremal*. We give a counterexample below.

$$\mathcal{F}_1 = \{\{c, d, e\}\}, Sh(\mathcal{F}_1) = \{\phi\}$$

$$\mathcal{F}_2 = \{\phi, \{a\}, \{b\}, \{a, b\}\}, Sh(\mathcal{F}_2) = \mathcal{F}_2$$

But, $Sh(\mathcal{F}_1 \cup \mathcal{F}_2) = \{\phi, \{a\}, \{b\}, \{a, b\}, \{c\}, \{d\}, \{e\}\}$ which is larger than the size of the family $\mathcal{F}_1 \cup \mathcal{F}_2$.

We will now see a characterization of *s-extremal* family based on *inclusion graph*. An *inclusion graph* $G_{\mathcal{F}}$ for a family $\mathcal{F}$ is defined as a graph whose vertices are all the sets in $\mathcal{F}$ and there is a directed edge from S to T, labelled with $j \in [n]$ when $T = S \cup \{j\}$.

Characterization of s-extremal Family Let $G_{\mathcal{F}}$ be the *inclusion graph* of the family $\mathcal{F}$. For a family $\mathcal{F}$ and a set $B \subseteq [n]$, define $\mathcal{F}(B) = \left\{I \subseteq [n] \setminus B \mid I + 2^B \subseteq \mathcal{F}\right\}$ where $I + 2^B = \{B' \cup I \mid B' \subseteq B\}$. Bollobás and Radcliffe (1995) shows the following characterization of *s-extremal* family.

Lemma 3.4.7 (Theorem 3 Bollobás and Radcliffe (1995)). *A family $\mathcal{F} \subseteq 2^{[n]}$ is s-extremal iff $G_{\mathcal{F}(B)}$ is connected for every $B \subseteq [n]$.*

To understand the above characterization, suppose $B = \phi$ then $\mathcal{F}(B) = \mathcal{F}$. The Lemma says that the *inclusion graph* of $\mathcal{F}$ should be connected. So if we find that *inclusion graph* of a family $\mathcal{F}$ is not connected, we immediately know that $\mathcal{F}$ is not *s-extremal*. Another characterization was given by Mészáros and Rónyai (2013) for a special class of family showing that *inclusion graph* must be a tree with distinct edge labels.

Lemma 3.4.8 (Mészáros and Rónyai (2013)). *A family $\mathcal{F}$ is* s-extremal *and of* VC-*dimension at most 1 iff $G_{\mathcal{F}}$ is a tree and all labels on the edges are different.*

Now we look at some open problems and conjectures for an *s-extremal* family. Mészáros and Rónyai (2013) conjectured the following about *s-extremal* families.

Conjecture 3.4.9 (Mészáros and Rónyai (2013)). *For an s-extremal family $\mathcal{F}$, there always exists a set $F \in \mathcal{F}$ such that $\mathcal{F} \setminus \{F\}$ is s-extremal.*

There is an equivalent formulation of this conjecture due to the fact that a family is *s-extremal* if and only if complement set of the family is *s-extremal* (see Bollobás and Radcliffe (1995)).

Conjecture 3.4.10 (Mészáros and Rónyai (2013))**.** *For an s-extremal family $\mathcal{F}$, there always exists a set $F \notin \mathcal{F}$ such that $\mathcal{F} \cup \{F\}$ is s-extremal.*

The conjecture is showed to be true for families such as *down-set family*, *s-extremal family with* VC*-dimension 1* etc. and proved for any family with VC-dimension at most 2 (see Mészáros and Rónyai (2013); Mészáros and Rónyai (2014)).

3.5 Complexity of Computing VC-dimension

In this section, we look at the hardness of computing VC-dimension for a given family $\mathcal{F}$. We saw several variants of VC-dimension in Chapter 2 and so we look at them individually. We first look at the variant due to Schafer (1996) (Definition 2.1.3).

VC-dimension Instance Let ℓ, m be two positive intergers. A circuit C with $\ell + m$ input gates, and a positive integer k. Question: Is $\mathsf{VC}(C) \geq k$?

Theorem 3.5.1 (Schafer (1996))**.** VC*-dimension is Σ_3^p-complete.*

We state the characterization to put the VC-dimension problem in Σ_3.

$$\begin{aligned} \mathsf{VC}(C) \geq k \Leftrightarrow & (\exists x_1, \ldots, x_k \in \Sigma^m) \\ & (\forall s \in \Sigma^k)(\exists i \in \Sigma^\ell) \\ & (\forall j \in \{1, \ldots, k\})[C(i, x_j) = s[j]] \end{aligned}$$

The problem is showed to be Σ_3^p-complete by reducing it to the standard Σ_3^p-complete problem QSAT$_3$. We refer the readers to Schafer (1996) for details of the proof.

Linial *et al.* (1991) first restricted representation of family or concept class to incidence matrix and termed the problem *Matrix* VC*-dimension* . They showed that *Matrix* VC*-dimension* can be solved in time $O(n^{\log n})$ where n is the number of entries in the matrix. Later Papadimitriou and Yannakakis (1996) defined the class LOGNP and showed that *Matrix* VC*-dimension* is LOGNP-complete. We will see the hardness of one more variant that is defined for a Single Boolean function in Chapter 5.

Approximability of VC-dimension Since the problem of deciding VC-dimension dimension is Σ_3^p-complete, the VC-dimension computing function is also complete for

$\mathsf{FP}^{\Sigma_3^p}$. Hence, Schafer (1996) looks at approximating the VC-dimension first, and show that approximating VC to a factor within $n^{1-\epsilon}$ is NP-hard for all $\epsilon > 0$. Thus we cannot approximate VC up to a constant factor in polynomial time, unless **P** = **NP**. On a similar line of work, Mossel and Umans (2002) shows that computing VC of a polynomial size circuit with n-inputs is Σ_3^p-hard to approximate within a factor of $2 - \epsilon$ and easy to approximate withing a factor of 2. They also show that the problem is approximable in $\mathcal{AM}$ (see Definition 2.3.2) within a factor of $2 - O(\frac{\log n}{\sqrt{n}})$ and $\mathcal{AM}$-hard to approximate within a factor of $n^{1-\epsilon}$ for all $\epsilon > 0$.

3.6 Other Applications of VC-dimension

In this section, we survey some more applications of VC-dimension .

Circuit Complexity Lower Bound As a first application of VC-dimension to circuit complexity, Koiran (1996) shows a $\Omega(n^{\frac{1}{4}})$ lower bound on the size of sigmoidal circuit computing a specific AC_2^0 function. For more details on circuit complexity bounds, we refer readers to the survey by Shpilka and Yehudayoff (2010).

One Way Randomized Communication Complexity Consider a randomized two-party communication protocol as defined by Yao (1979). Alice holds an input $x \in \{0,1\}^n$ and Bob holds an input $y \in \{0,1\}^n$ and they want to compute a function $f(x,y)$. In a one-round protocol, Alice will send a single message(depending on her input and random coins) to Bob who should be then able to compute the function $f(x,y)$. The communication complexity of this protocol or the number of bits of communication is denoted by $R^{A\to B}(f)$. Kremer *et al.* (1995) shows that $R^{A\to B}$ is lower bounded by VC-dimension of the function $f(x,y)$. Notice that the family corresponding to function $f(x,y)$ is the one defined by Schafer (1996) (see Definition 2.1.3). We refer the readers to textbook by Kushilevitz and Nisan (1996) and Rao and Yehudayoff (2020) for details about communication complexity.

VC-dimension and Disjointness Problem The disjointness problem - where Alice and Bob are given two subsets of $[n]$, and they need to check whether their sets intersect, with the least amount of communication. The deterministic and randomized communication complexity of the disjointness problem is known to be $\Theta(n)$ but if the sets drawn are from some restricted set system, then the communication complexity is much lower.

Bhattacharya *et al.* (2020) construct a set system of VC-dimension d and show that, when sets are drawn from this set system, deterministic and randomized communication complexity is $\Theta(d \log(\frac{n}{d}))$. The intersection problem is- given two subsets of $[n]$, find their intersection. They show that there exists a set system of VC-dimension d such that deterministic and randomized communication complexities of the set intersection problem is $\Theta(d \log(\frac{n}{d}))$.

CHAPTER 4

Alternation, k-fold Union and VC-dimension

There have been several works exploring tight VC-dimension bounds of various families of Boolean functions due to their application in proving sample complexity bounds of learning algorithms. We saw in Chapter 3, that VC-dimension yields matching lower bound (Ehrenfeucht *et al.* (1989)) and upper bound (Blumer *et al.* (1989)) for the number of samples. In this chapter, we start by showing VC-dimension bounds for *k-slice* family, *Threshold* family and family of Boolean functions with *alternation* at most 1 (denoted by $\mathcal{F}_1$). We show that $\mathsf{VC}(\mathcal{F}_1) = \binom{n}{n/2} + 1$. We parameterize family of non-monotone functions with *alternation* as a measure of non-monotonicity and show VC-dimension bounds in terms of this parameter (i.e. alternation). For *alternation* k at most $\sqrt{n}$ we give an asymptotically matching lower and upper bound. As an application, we show tightness of VC-dimension bounds for k-fold union, by explicitly constructing a family $\mathcal{F}$ of subsets of $\{0,1\}^n$ such that k-fold union of the family, $\mathcal{F}^{k\cup} = \left\{\bigcup_{i=1}^{k} F_i \mid F_i \in \mathcal{F}\right\}$ must have VC-dimension at least $\Omega(dk)$ and that this bound holds even when the union is over disjoint sets from $\mathcal{F}$. For a k-union family Blumer *et al.* (1989) showed an upper bound of $O(dk \log k)$ and the bound was shown to be tight by Eisenstat and Angluin (2007) by constructing a geometric family with VC-dimension at least $\Omega(dk \log k)$. We gave a non-geometric and explicit construction of a k-union family. Finally, we show *s-extremal* properties of family of *Monotone* Boolean and show that Mészáros and Rónyai (2013) conjecture is true for the family of Monotone Boolean function.

We also look at the notion of Subspace VC-dimension and show bounds for several families such as *Monotone family, Symmetric family* and *Family of Parity*.

4.1 VC-dimension for functions with alternation 1

As a warm-up towards later sections, in this section, we describe VC-dimension bounds for families of functions related to alternation and monotonicity. As mentioned in the literature survey, Procaccia and Rosenschein (2006) computes the VC-dimension of family of monotone Boolean functions (denoted by $\mathcal{M}$) as : $\mathsf{VC}(\mathcal{M}) = \binom{n}{n/2}$. We see VC-dimension bounds of few more related families.

k-slice function A function $f : \{0,1\}^n \rightarrow \{0,1\}$ is said to be k-slice function if $f(x) = 0$ if for every $x \in \{0,1\}^n$, $\sum_{i=1}^{n} x_i < k$ and 1 if $\sum_{i=1}^{n} x_i > k$. Define $\mathcal{M}^*$ to be the family of all $slice$-functions. We compute the VC-dimension($\mathcal{M}^*$).

Proposition 4.1.1. $\mathsf{VC}(\mathcal{M}^*) = \binom{n}{n/2}$

Proof. To show the lower bound, we demonstrate how to shatter the following set :
$S = \{x \in \{0,1\}^n \mid \sum_{i=1}^{n} x_i = n/2\}$ where x_i represents i^{th} bit of x and

$$|S| = \binom{n}{n/2}.$$

We can see that, $\forall S' \subseteq S$, F such that $F \cap S = S'$ is given by the characteristic function $f = \bigvee_{z \in S'} \bigwedge_{z_i = 1} x_i$. Observe that f is $\frac{n}{2}$-$slice$ function. To see the upper bound, observe that every slice function is a monotone function and hence slice family is a subset of family of monotone Boolean functions. Hence $\mathsf{VC}(\mathcal{M}^*) \leq \mathsf{VC}(\mathcal{M}) \leq \binom{n}{n/2}$. □

Threshold Family We define T_n as class of THRESHOLD *functions*.

$$T_n = \left\{ \mathrm{TH}_n^k \mid k \in \{0,1,\ldots,n\} \right\}$$

where

$$\mathrm{TH}_n^k = \begin{cases} 1, & if \sum_{i=1}^{n} x_i \geq k \\ 0, & otherwise \end{cases}$$

We need the following proposition before proving the bound for *Threshold* family.

Proposition 4.1.2. *For $t > 0$, let $f_1, f_2 \ldots, f_t : \{0,1\}^n \rightarrow \{0,1\}$ be t Boolean functions such that $f_1 \implies f_2 \implies \ldots \implies f_t$. Let $\mathcal{F}$ be the family containing these t functions. Then* $\mathsf{VC}(\mathcal{F}) = 1$.

Proof. Suppose , it was the case that $\mathsf{VC}(\mathcal{F}) \geq 2$ and shattered set is $S = \{w_1, w_2\}$. Then there exists, some $A \cup \{w_1\} \in \mathcal{F}$ and $B \cup \{w_2\} \in \mathcal{F}$. Now, either $A \cup \{w_1\} \subseteq B \cup \{w_2\}$ or vice-versa. In either case, it contradicts the shattering of $S = \{w_1, w_2\}$ as any one of $\{w_1\}$ or $\{w_2\}$ can be obtained but not both while shattering. To see the lower bound, consider a set $S = \{w\}$ and two functions f_i and f_j such that $f_i \implies f_j$ and $w \in f_j \setminus f_i$ (without loss of generality). The set S can be shattered by obtaining empty set from f_i and full set from f_j. □

Proposition 4.1.3. $\mathsf{VC}(T_n) = 1$

Proof. We first observe that the family T_n has implication property i.e. for all $0 \leq i \leq n-1$, $\text{TH}_n^{i+1} \implies \text{TH}_n^i$. Using Proposition 4.1.2 we have that $\mathsf{VC}(T_n) = 1$. □

alt-1 family Now consider the family of functions where each function has alternation at most 1. Define

$$\mathcal{F}_1 = \{f \mid \text{either } f \text{ or } \neg f \text{ is monotone}\}$$

We compute the VC-dimension of this family exactly.

Theorem 4.1.4. $\mathsf{VC}(\mathcal{F}_1) = \binom{n}{\lfloor n/2 \rfloor} + 1$

Proof. Lower Bound: We show the lower bound by shattering a set $S \subseteq U$ of cardinality $\binom{n}{\lfloor n/2 \rfloor} + 1$. The shattered set $S = \{x \in \{0,1\}^n \mid \sum_{i=1}^n x_i = \lfloor n/2 \rfloor\} \bigcup \{w\}$ where w is any arbitrary point in $\{0,1\}^n$ such that $wt(w) < \lfloor n/2 \rfloor$. Let $S' \subseteq S$. We need to give an $F \in \mathcal{F}$ such that $F \cap S = S'$. We consider the following cases:

- **Case 1:** $S' \subseteq \{x \in \{0,1\}^n \mid \sum_{i=1}^n x_i = n/2\}$. F such that $F \cap S = S'$ is given by the characteristic function $f = \bigvee_{z \in S'} \bigwedge_{z_i=1} x_i$. In this case S' forms an *antichain* and hence we describe the f which is *True* on this set. See Figure 4.3

- **Case 2:** $S' = X \cup \{w\}$ where $\forall x \in X, \sum_{i=1}^n x_i = n/2$. F such that $F \cap S = S'$ is given by the characteristic function $f = \bigwedge_{z \in S'} \bigvee_{z_i=1} \overline{x_i}$. Observe that negation of this function is a monotone function. The set S' may or may not be an *antichain*. Hence, we use negation of a monotone function to obtain the set. See Figure 4.4 . We also note here that, the function defined will not change as long as $wt(w) < n/2$.

Upper Bound: Suppose $\mathcal{F}_1$ shatters a set S such that $|S| \geq \binom{n}{n/2} + 2$. We first obtain certain properties that S cannot have through the following lemma. We define two properties first.

A set H is said to have **parallel chain property**, when there are elements $p, p', q, q' \in H$ such that all of them are distinct and $p \preceq p'$ and $q \preceq q'$. A set H is said to have **triplet**

chain property, when there are elements $p, q, r \in S$, all distinct, such that $p \preceq q \preceq r$. See Figure 4.1 for a pictorial view.

Lemma 4.1.5. *$\mathcal{F}_1$ cannot shatter a set S if it has either of the parallel chain or triplet chain properties.*

Proof. For parallel chain, without loss of generality let us suppose that $p \preceq p'$ and $q \preceq q'$ and all the elements are in S. Suppose $S' = \{p', q\}$. Now, any monotone function which evaluates to 1 on q will also evaluate to 1 on q'. Similarly, any monotonically decreasing function which evaluates to 1 on p', will also evaluate to 1 on p. Thus, there does not exist a function $g \in \mathcal{F}_1$ such that $g \cap S = S'$. See Figure 4.2.

For triplet chain, suppose $p, q, r \in S$ such that $p \preceq q \preceq r$. Consider the set $S' = \{p, r\}$. We observe that there does not exist a function $f \in \mathcal{F}_1$ such that it is true on p and r and evaluates to false on q. Hence S cannot be shattered. See Figure 4.2

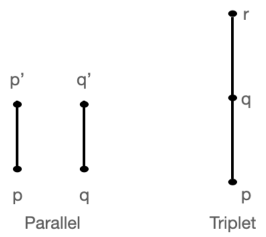

Figure 4.1: Parallel and Triplet Chain

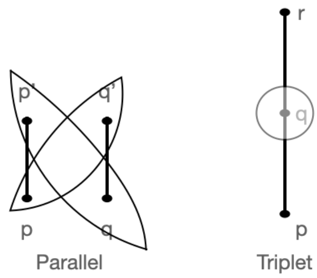

Figure 4.2: Shattering Parallel and Triplet Chain

□

Now we claim that set S which is shattered must be a disjoint union of at most 2 maximal antichains. To see this, suppose $S = S_1 \uplus S_2 \uplus S_3$ such that the sets are maximal antichain. Without loss of generality, consider an element $p \in S_1$. Now there exists a point $q \in S_2$ such that p and q are comparable (otherwise they will be in the same set). Now if this is the case then neither p nor q can be related to any point in S_3 as it will either form a chain length of 3 i.e. $p \prec q \prec r$(triplet chain) or parallel chains because of which the set S cannot be shattered (see Lemma 4.1.5). Now if p and q are not comparable to any element in S_3 then we can have a larger antichain by including either p or q in the set S_3 which contradicts the maximality of antichain set S_3. The same argument can be given for each of the sets. Hence set S can be disjoint union of at most 2 maximal antichains.

We now conclude that there must be $S = S_1 \uplus S_2$ where S_1, S_2 are maximal antichains i.e. no elements from S_1 can be put into S_2 and vice versa. This is because $|S| \geq \binom{n}{n/2}+2$ and the largest antichain set can be of size $\binom{n}{n/2}$. Using Lemma 4.1.5 again, we have that S does not have either parallel chain or triplet chain property. But due to this, we can contradict the maximality of set S_1 in the following way. Consider two elements $p', q' \in S_2$. If S didn't have any parallel or triplet chain, then it means that p', q' don't have any comparable elements in S_1 and hence they can be brought inside S_1 which contradicts the fact that S_1 was maximal. In a similar way, we can argue that S_2 was not maximal if there are no parallel or triplet chain property in S. Thus, S must have a parallel or triplet chain property.

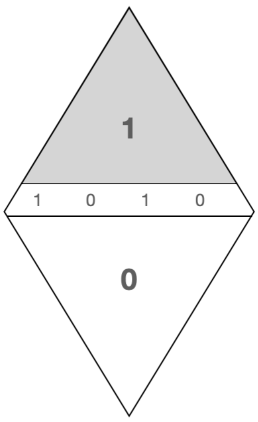

Figure 4.3: Monotone function

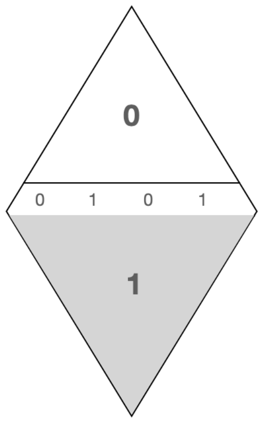

Figure 4.4: Negation of Monotone function

□

4.2 Bounds for VC-dimension in terms of $\mathsf{alt}(f)$

In this section, we derive VC-dimension bounds for families of Boolean functions, parameterized by the maximum alternation of functions in the family.

Upper Bound : We need the following known lemma.

Lemma 4.2.1 (Blais *et al.* (2015))**.** *Let* $f : \{0,1\}^n \to \{0,1\}$ *. Then there exists*

$k = \mathsf{alt}(f)$ *monotone Boolean functions* $g_1, \ldots, g_k$ *each from* $\{0,1\}^n$ *to* $\{0,1\}$ *such that*

$$f(x) = \begin{cases} \oplus_{i=1}^{k} g_i & , \textit{if } f(0^n) = 0 \\ \neg \oplus_{i=1}^{k} g_i & , \textit{if } f(0^n) = 1 \end{cases}$$

We use this lemma to establish the following upper bound:

Theorem 4.2.2. *Let* $k > 1$. *If* $\mathcal{F}_k$ *is the family of Boolean functions* f *such that* $\mathsf{alt}(f) \leq k$. *Then,* $\mathsf{VC}(\mathcal{F}_k) \leq O\left(k\binom{n}{n/2}\log k\right)$.

Proof. We apply Lemma 4.2.1 to conclude that $\mathcal{F}_k$ can be equivalently written as

$$\mathcal{F}_k = \left\{\oplus_{i=1}^{k} f_i \mid f_i \in \mathcal{M} \text{ and } \oplus_{i=1}^{k} f_i(0^n) = 0\right\} \bigcup \left\{\neg \oplus_{i=1}^{k} f_i \mid f_i \in \mathcal{M} \text{ and } \neg \oplus_{i=1}^{k} f_i(0^n) = 1\right\}$$

where $\mathcal{M}$ is the family of Monotone Boolean functions. We look at a family

$$\mathcal{G} = \left\{f \oplus g \mid f = \oplus_{i=1}^{k} f_i, f_i \in \mathcal{M}, g(x) = 1 \text{ if } f(0^n) = 1 \text{ and } g(x) = 0 \text{ if } f(0^n) = 0\right\}$$

Observe that, $\mathcal{F}_k \subseteq \mathcal{G}$ and hence $\mathsf{VC}(\mathcal{F}_k) \leq \mathsf{VC}(\mathcal{G})$. We turn to the following lemma to show an upper bound in general for such constructed families. Given k classes of n-bit Boolean functions $\mathcal{F}_1, \mathcal{F}_2 \ldots, \mathcal{F}_k$, and a fixed Boolean function $f : \{0,1\}^k \to \{0,1\}$. We define

$$\mathcal{F}(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k) = \{f(f_1(), ..., f_k()) \mid f_i \in \mathcal{F}_i, i \in [k]\}$$

where same inputs are being fed to all the f_i's. To understand this family better, consider a function $g : \{0,1\}^n \to \{0,1\}$ in the family $\mathcal{F}(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k)$. We have, $\forall x \in \{0,1\}^n$, $g(x) = f(f_1(x), f_2(x), \ldots, f_k(x))$. We have the following lemma,

Lemma 4.2.3 (Blumer *et al.* (1989); Haussler and Welzl (1987); Sontag (1998))**.** *Let* $d = \max_{i \in [k]}(\mathsf{VC}(\mathcal{F}_i))$. $\mathsf{VC}(\mathcal{F}(\mathcal{F}_1, \ldots, \mathcal{F}_k)) \leq O(dk \log k)$

Observe that, each member of the family $\mathcal{G}$ is parity ($\oplus$) of $k+1$ monotone Boolean functions. Thus applying this lemma, we obtain for the above family $\mathsf{VC}(\mathcal{F}_k) \leq O\left(k\binom{n}{n/2}\log k\right)$. ☐

For small values of *alternation* k, the above bound is almost in range with the lower bound of VC-dimension of such families as we will see in the later part of this chapter. However, for large values k, the bound is not sharp. For $k = n$, the upper bound is approximately $O(n\binom{n}{n/2}\log n)$ which is much larger than the entire universe size 2^n. We now will see another approach to further improve the upper bound in the below theorem.

Theorem 4.2.4. *Let $k > 1$. If $\mathcal{F}_k$ is the family of Boolean functions f such that $\mathsf{alt}(f) \leq k$. Then, $\mathsf{VC}(\mathcal{F}_k) \leq O\left(k\binom{n}{n/2}\right)$.*

Proof. Similar to Theorem 4.2.2, we apply Lemma 4.2.1 to conclude that $\mathcal{F}_k$ can be equivalently written as $\mathcal{F}_k = \left\{(\neg \oplus \ \ or\ ) \oplus_{i=1}^{k} f_i \mid f_i \in \mathcal{M}\right\}$ where $\mathcal{M}$ is the family of Monotone Boolean functions. We look at a family,

$$\mathcal{G} = \left\{f \oplus g \mid f = \oplus_{i=1}^{k} f_i, f_i \in \mathcal{M}, g = const\right\}$$

where $g(x) = 1$ if $f(0^n) = 1$ and $g(x) = 0$ if $f(0^n) = 0$. Observe that, $\mathcal{F}_k \subseteq \mathcal{G}$ and hence $\mathsf{VC}(\mathcal{F}_k) \leq \mathsf{VC}(\mathcal{G})$. The idea is simple counting : we have $|\mathcal{G}| \leq |\mathcal{M}|^{k+1}$. We know that $\mathsf{VC}(\mathcal{G}) \leq \log(|\mathcal{G}|)$, which gives us $\mathsf{VC}(\mathcal{G}) \leq (k+1)\log(|\mathcal{M}|)$. This bound is the *Dedekind's number* and we use the following bound due to Kleitman and Markowsky (1975) : $\log(|\mathcal{M}|) \leq \binom{n}{n/2}\left(1 + O(\frac{\log n}{n})\right)$. This gives, $\mathsf{VC}(\mathcal{F}_k) \leq (k+1)\binom{n}{n/2}\left(1 + O\left(\frac{\log n}{n}\right)\right)$ Hence, we have $\mathsf{VC}(\mathcal{F}_k) \leq O(k\binom{n}{n/2})$. □

Lower Bound : Now we turn to the lower bound. We have seen the VC-dimension bound for $k = 1$ (Theorem 4.1.4). Before generalizing to any k, we look at $k = 2$. Define the family $\mathcal{F}_2$ as the family of Boolean functions with *alternation* at most 2.

$$\mathcal{F}_2 = \{f : \{0,1\}^n \to \{0,1\} \mid \mathsf{alt}(f) \leq 2)\}$$

As a first attempt, we apply Theorem 3.4.3, to obtain a bound on the VC-dimension of the family $\mathcal{F}_2$. Using Lemma 4.2.1, we have that $\mathcal{F}_2$ can be equivalently written as $\mathcal{M}\triangle\mathcal{M}$ where $\mathcal{M}$ denotes family of Monotone Boolean functions. Suppose $\mathsf{VC}(\mathcal{F}_2) \leq d'$, then from Theorem 3.4.3 we have $|\mathcal{M}| \leq 2\sum_{i=0}^{d'/2}\binom{2^n}{i}$. We also know that $|\mathcal{M}| \geq$

$2^{\binom{n}{n/2}}$ (no. of monotone Boolean functions). We have,

$$\begin{aligned}|\mathcal{M}| &\leq 2\sum_{i=0}^{d'/2}\binom{2^n}{i}\\ &\leq 2.\left(\frac{2e2^n}{d'}\right)^{d'/2}, \text{ using } \binom{n}{k} \leq \left(\frac{en}{k}\right)^k\\ \Rightarrow \log(|\mathcal{M}|) &\leq d'n + 1\end{aligned}$$

We know that $\log(|\mathcal{M}|) \geq \binom{n}{n/2}$. From this we obtain that $\mathsf{VC}(\mathcal{F}_2) \geq \frac{1}{n}\left(\binom{n}{n/2} - 1\right)$. However, using the fact that for any $k \geq 1$, the family $\mathcal{F}_k$ also includes the set of monotone functions $\mathcal{M}$, the VC-dimension$(\mathcal{F}_k) \geq \mathsf{VC}$-dimension$(\mathcal{M})$. Hence VC-dimension$(\mathcal{F}_k) \geq \binom{n}{n/2}$. Thus this method gives much weaker lower bound. We will further improve this in the below theorem:

Theorem 4.2.5. *Let $k > 1$. If $\mathcal{F}_k$ is the family of Boolean functions f such that $\mathsf{alt}(f) \leq k$. Then, $\mathsf{VC}(\mathcal{F}_k) \geq \sum_{i=n/2-k/2}^{n/2+k/2}\binom{n}{i}$*

Proof. We shatter the set

$$S = \left\{x \in \{0,1\}^n \mid n/2 - k/2 \leq \sum_{i=1}^{n} x_i \leq n/2 + k/2\right\}$$

To obtain any $S' \subseteq S$, we give $f \in \mathcal{F}_k$ such that $f(x) = 1$ whenever $x \in S'$ and $f(x) = 0$ otherwise. It remains to show that $\mathsf{alt}(f) \leq k$. Since $wt(x)$ for $x \in S'$ can only be in the range $[n/2-k/2, n/2+k/2]$, any chain in this part will have alternation at most k and in the remaining part 0. Hence we obtain all $S' \subseteq S$. We show the shattered set pictorially in Figure 4.5.

□

VC-dimension of Read-once functions: A Boolean function is said to be *read-once* if there is a Formula (a Boolean circuit where every gate has fanout at most 1) computing the function f such that every variable appears only once in the formula. Monotone read-once functions have no negations in the formula computing the function. The following lemma motivates the study of VC-dimension under composition between read-once functions and functions with alternation at most 1.

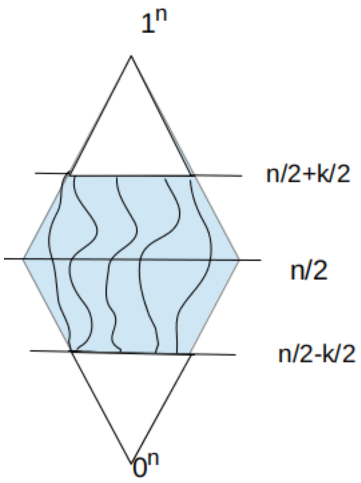

Figure 4.5: Shattered set for alt-k family

Lemma 4.2.6. *Let* $f : \{0,1\}^n \rightarrow \{0,1\}$ *such that* $\mathsf{alt}(f) \leq k$*. Then* $f = g(f_1, f_2, \ldots, f_k)$ *where* g *is a monotone-Read Once formula and* $f_i \in \mathcal{F}_1$ *where* $\mathcal{F}_1$ *is the family of functions with alternation at most* 1*.*

Proof. Let $f : \{0,1\}^n \rightarrow \{0,1\}$ such that $\mathsf{alt}(f) \leq k$. We first show this for case when k is even.

Due to the alternation characterization in Lemma 4.2.1, we have

$$f = \bigoplus_{i=1}^{k} f_i = \bigoplus_{i=1}^{k/2} (f_{2i-1} \oplus f_{2i})$$
$$= \bigoplus_{i=1}^{k/2} (\neg\, f_{2i-1} \wedge f_{2i}) \vee (f_{2i-1} \wedge \neg\, f_{2i})$$

It can be observed from the construction in Blais *et al.* (2015) that $f_i \rightarrow f_{i+1}$. Using this we have that $f_i \wedge \neg f_{i+1}$ always evaluates to 0. Thus,

$$f = \bigvee_{i=1}^{k/2} (\neg\, f_{2i-1} \wedge f_{2i})$$

Now we show the case when k is odd.

Due to the alternation characterization in Lemma 4.2.1, we have

$$
\begin{aligned}
f = \bigoplus_{i=1}^{k} f_i &= (\bigoplus_{i=1}^{(k-1)/2} (f_{2i-1} \oplus f_{2i})) \bigoplus f_k \\
&= (\bigoplus_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i}) \vee (f_{2i-1} \wedge \neg f_{2i})) \bigoplus f_k \\
&= (\bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i})) \bigoplus f_k \quad (\text{ using } f_i \to f_{i+1}) \\
&= \left((\bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i})) \wedge \neg f_k \right) \bigvee \left((\neg \bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i})) \wedge f_k \right) \\
&= \left((\bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i} \wedge \neg f_k)) \right) \bigvee \left((\neg \bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i})) \wedge f_k \right) \\
&= (\neg \bigvee_{i=1}^{(k-1)/2} (\neg f_{2i-1} \wedge f_{2i})) \wedge f_k \quad (\text{ using } f_{2i} \wedge \neg f_k = 0) \\
&= f_k \wedge \bigwedge_{i=1}^{(k-1)/2} (f_{2i-1} \vee \neg f_{2i}) \quad (\text{ using DeMorgan's Law})
\end{aligned}
$$

Hence, we obtained $f = \bigvee_{i=1}^{k/2} (\neg f_{2i-1} \wedge f_{2i})$ when k is even and $f = f_k \wedge \bigwedge_{i=1}^{\frac{k-1}{2}} (f_{2i-1} \vee \neg f_{2i})$ when k is odd. Observe that in the formula each f_i appears only once and no negation in the formula is involved. Since f_i's are monotone, hence $f_i, \neg f_i \in \mathcal{F}_1$. □

This gives us a motivation to study VC-dimension of monotone read-once functions which could potentially be applied to improve the bound of alt-k family.
Let $\mathcal{R} = \{f : \{0,1\}^n \to \{0,1\} \mid f \text{ is monotone read-once}\}$. We prove the following

Proposition 4.2.7. $n \leq \mathsf{VC}(\mathcal{R}) \leq O(n \log n)$

Proof. To see the lower bound a subclass of this family was studied in Natschläger and Schmitt (1996) known as Monotone monomial whose VC-dimension is n. To see the upper bound we will count all possible monotone read-once formulas on n-variables.

Claim 4.2.8. $|\mathcal{R}| = n!\binom{2n}{n}$

Proof. A formula is rooted binary tree such that all the internal nodes are $\wedge, \vee$ or$\neg$. The number of full binary trees[1] with $n+1$ labeled leaves is $\prod_{i=1}^{n}(2i-1)$ (see Barnett *et al.* (2010)). But in our case we also have to consider the two choices $\{\wedge, \vee\}$ for each internal nodes. Hence we get $|\mathcal{R}| = 2^n . \prod_{i=1}^{n}(2i-1)$ which gives $|\mathcal{R}| = n!\binom{2n}{n}$. □

Using the above claim we obtain, $\mathsf{VC}(\mathcal{R}) \leq \log |\mathcal{R}| = O(n \log n)$. □

We saw VC-dimension bounds of several classes of Boolean functions such as *Read-once, Monotone, Monomials* etc. We also saw a general upper bound for VC-dimension of k-closure of families and we improved the bound using decomposition of a non-monotone function into monotone functions. But, the improvement is given for a particular class of Boolean functions and hence, there are avenues to improve for more such k-closures. We will now see an application of the VC-dimension of the family $\mathcal{F}_k$.

4.3 Application to VC-dimension of k-fold Union

In this section, we show VC-dimension bound for a non-geometric family which is a k-union. We also remark in the end that even if we restrict our family to have only disjoint union of k functions, we obtain a VC-dimension bound of $\Omega(dk)$.

Lemma 4.3.1. *Let* $\mathcal{F}_{2k} = \{f : \{0,1\}^n \to \{0,1\} \mid \mathsf{alt}(f) \leq 2k\}$. *Then this family is same*[2] *as* $\mathcal{G} = \left\{\bigcup_{i=1}^{k} g_i \mid g_i : \{0,1\}^n \to \{0,1\}, \mathsf{alt}(g_i) \leq 2\right\}$

Proof. Let $f \in \mathcal{F}_{2k}$. Due to alternation characterization described in Lemma 4.2.1 we have, $f = \bigoplus_{i=1}^{2k} f_i = \bigoplus_{i=1}^{k}(\neg f_{2i-1} \wedge f_{2i}) \vee (f_{2i-1} \wedge \neg f_{2i})$. It can be observed from the construction of Blais *et al.* (2015), that $f_i \to f_{i+1}$. Now using this fact we obtain $f = \bigvee_{i=1}^{k}(\neg f_{2i-1} \wedge f_{2i}) = \bigvee_{i=1}^{k} g_i$ such that $\mathsf{alt}(g_i) \leq 2$. In fact something stronger is true - we argue that f is the disjoint union of k sets.

Claim 4.3.2. *Let* $f : \{0,1\}^n \to \{0,1\}$ *and* $\mathsf{alt}(f) \leq 2k$. *Then* $f = \uplus_{i=1}^{k} g_i$ *i.e. disjoint union of* k *sets.*

Proof. From the proof of Lemma 4.2.6 we have, $f = \bigvee_{i=1}^{k}(\neg f_{2i-1} \wedge f_{2i})$. Let $g_i = (\neg f_{2i-1} \wedge f_{2i})$. We will show that g_1 and g_2 are disjoint which will be applicable for all

[1] A binary tree where each node has either 0 or 2 children.

[2] Under the interpretation of the sets in the system as $f^{-1}(1)$ for Boolean functions, $\vee$ and $\cup$ are used interchangeably.

$i, j \in [k]$.

Suppose for $x \in \{0,1\}^n$, $g_1(x) = 1 \Rightarrow f_1(x) = 0, f_2(x) = 1$. To show a contradiction, suppose $g_2(x) = 1$ which implies that $f_3(x) = 0, f_4(x) = 1$. We also have $f_i \to f_{i+1}$. We obtained $f_2(x) = 1$ but $f_3(x) = 0$ which contradicts $f_i \to f_{i+1}$. Hence g_1 and g_2 are disjoint. □

Now we need to argue the reverse direction. We have a Boolean function $g : \{0,1\}^n \to \{0,1\}$, $g = \bigvee g_i$ where $g_i : \{0,1\}^n \to \{0,1\}$ and $\mathsf{alt}(g_i) \leq 2$. We need to show that $\mathsf{alt}(g) \leq 2k$. We use the property that $\mathsf{alt}(g_1 \vee g_2) \leq \mathsf{alt}(g_1) + \mathsf{alt}(g_2)$ iteratively to conclude that $\mathsf{alt}(g) \leq 2k$ which we prove in the claim below.

Claim 4.3.3. *Let* $g_1 : \{0,1\}^n \to \{0,1\}$ *and* $g_2 : \{0,1\}^n \to \{0,1\}$ *with* $\mathsf{alt}(g_1) = k_1$ *and* $\mathsf{alt}(g_2) = k_2$. *Then* $\mathsf{alt}(g_1 \vee g_2) \leq k_1 + k_2$.

Proof. Suppose on the Boolean hypercube, $x, y \in \{0,1\}^n, x \preceq y$ and they differ exactly at one bit. Then we will say (x, y) forms an edge. Now we say (x, y) is monochromatic when function $f(x) = f(y)$ and bichromatic when $f(x) \neq f(y)$. Clearly for any function $f : \{0,1\}^n \to \{0,1\}$ with $\mathsf{alt}(f) \leq k$, there are at most k-bichromatic edges on any chain and vice versa. Now we count the number of bichromatic edges on any chain of $g_1 \vee g_2$. Suppose on some chain in g_1, there is a monochromatic edge with $g_1(x) = g_1(y) = 1$ then this edge will stay monochromatic even after OR with g_2. But if the edge (x, y) is monochromatic and $g_1(x) = g_1(y) = 0$ then they can become bichromatic if they are ORed with a bichromatic edge. Hence the number of bichromatic edges on any chain of $g_1 \vee g_2$ can at most be increased by number of bichromatic edges on any chain in g_2. Hence $\mathsf{alt}(g_1 \vee g_2) \leq k_1 + k_2$. □

□

Theorem 4.3.4. *Let* $\mathcal{F}_2 = \{f : \{0,1\}^n \to \{0,1\} \mid \mathsf{alt}(f) \leq 2\}$. *Consider the family* $\mathcal{F}^{k\cup} = \left\{\bigcup_{i=1}^{k} f_i \mid f_i \in \mathcal{F}_2\right\}$. *For* $k \leq \Theta(\sqrt{n})$, *we have* $\mathsf{VC}(\mathcal{F}^{k\cup}) = \Theta(k\binom{n}{n/2})$.

Proof. From Lemma 4.3.1, we have that the family $\mathcal{F}^{k\cup}$ can alternately be represented as parity-composition of a family of monotone Boolean functions. So $\mathcal{F}^{k\cup} = \mathcal{F}_{2k}$. We conclude using Theorem 4.2.4 that $\mathsf{VC}(\mathcal{F}^{k\cup}) \leq O(k\binom{n}{n/2})$. From Theorem 4.2.5 we have: $\mathsf{VC}(\mathcal{F}^{k\cup}) \geq \sum_{i=n/2-k}^{n/2+k} \binom{n}{i}$.

$$\sum_{i=n/2-k}^{n/2+k} \binom{n}{i} = \binom{n}{n/2} + 2\sum_{i=1}^{k} \binom{n}{n/2+i}$$

Now we use the following bounds due to Spencer and Florescu (2014). When i is $o(n^{2/3})$, each summand in the second term is equal to $\binom{n}{n/2}e^{-\frac{2i^2}{n}+O(i^3n^{-2})}$. Hence for $i \leq c\sqrt{n}$, we obtain $\binom{n}{n/2+i} = c_1\binom{n}{n/2}$, $c_1 > 0$. Hence we obtain the lower bound as

$$\sum_{i=n/2-k}^{n/2+k} \binom{n}{i} = \Omega(k\binom{n}{n/2})$$

□

4.4 Extremal Properties of family of Monotone functions

Recall the discussion on shattering-extremality and related results in Section 3.4.2 . We now show that $\mathcal{M}$ (family of Monotone Boolean functions) exhibits *shattering-extremal* property.

Proposition 4.4.1. $\mathcal{M}$ *is* s-extremal *i.e.* $|Sh(\mathcal{M})| = |\mathcal{M}|$.

Proof. We show a bijection between the two sets $Sh(\mathcal{M})$ and $\mathcal{M}$. We first claim that $Sh(\mathcal{M})$ is exactly the set of antichains in the Boolean hypercube. To see this, we first argue that $Sh(\mathcal{M})$ does not contain non-antichains. For the sake of contradiction, let $S \in Sh(\mathcal{M})$ such that $p, q \in S$ such that $p \preceq q$. Consider the subset $S' = \{p\} \subseteq S$. Observe that, for any $f \in \mathcal{M}$, $f(p) = 1 \implies f(q) = 1$. Hence $f \cap S \neq S'$. This contradicts the fact that S was shattered by the family $\mathcal{M}$. The family $\mathcal{M}$ can shatter all sets which are an antichains (see Procaccia and Rosenschein (2006)). Thus, $Sh(\mathcal{M})$ contains all the antichains of the Boolean hypercube. Hence the claim.

Now it suffices to argue that there is a bijection between set of monotone functions $\mathcal{M}$ and set of antichains in the Boolean hypercube. This uses an idea from Kleitman and Markowsky (1975) and we describe it below.

Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone Boolean function. We show how to obtain the set $A \subseteq \{0,1\}^n$ which is the antichain corresponding to the function f. First, obtain the minimal monotone-DNF formula for f. The minimal monotone-DNF formula for a monotone function is unique (see Crama and Hammer (2011)). For each positive literal appearing in a term of the formula, the corresponding bit will be set to 1 and the rest as 0. For example, a term $x_1 \wedge x_2$ in minimal monotone-DNF formula will correspond to a string $w = 110\ldots0$. Thus, each term in the formula will correspond to an element in A. Observe that, in this construction if the monotone-DNF formula is not minimal, then the set A will not be an antichain.

We will now argue the *one-one* relation between the function f and the corresponding antichain A. Consider two monotone Boolean functions f_i, f_j such that $f_i \neq f_j$. Then, their minimal monotone-DNF formula will differ in at least 1 term. Hence, their corresponding antichain A_i, A_j will also differ as per our construction.

Using a similar reverse construction, for every antichain, we can obtain the minimal monotone-DNF formula of the monotone Boolean function.

□

Mészáros and Rónyai (2013) conjectured the following about *s-extremal* families. We restate it here and show that the conjecture holds for family of *Monotone* Boolean functions.

Conjecture 4.4.2 (Mészáros and Rónyai (2013))**.** *For an s-extremal family $\mathcal{F}$, there always exists a set $F \in \mathcal{F}$ such that $\mathcal{F} \setminus \{F\}$ is s-extremal.*

There is an equivalent formulation of this conjecture due to the fact that a family is *s-extremal* if and only if complement set of the family is *s-extremal* Bollobás and Radcliffe (1995).

Conjecture 4.4.3 (Mészáros and Rónyai (2013))**.** *For an s-extremal family $\mathcal{F}$, there always exists a set $F \notin \mathcal{F}$ such that $\mathcal{F} \cup \{F\}$ is s-extremal.*

The conjecture has been showed to be true for families such as *down-set family*, *s-extremal family with* VC*-dimension 1* etc. and proved for any family with VC-dimension at most 2 (see Mészáros and Rónyai (2013); Mészáros and Rónyai (2014)). We now show that this conjecture holds for family of *Monotone Boolean functions* , whose VC-dimension is very high. We demonstrate a Boolean function that can be removed and another which can be added such that the resulting family is still *s-extremal*.

Proposition 4.4.4. *Let $\mathcal{M}$ be the family of Monotone Boolean functions and $g : \{0,1\}^n \to \{0,1\}$ is constant 1 function. Then $\mathcal{M} \setminus \{g\}$ is* s-extremal.

Proof. Upon the removal of the constant 1 function ($\forall x \in \{0,1\}^n,\ g(x) = 1$), the family will shatter one less set which is $\{0^n\}$. It is because any *monotone* function which is *True* on 0^n input is also *True* on every other input by definition. This function is clearly constant 1 function. Since we removed this, we cannot shatter $\{0^n\}$ anymore. All the other shattered sets are intact. □

Proposition 4.4.5. *Let $\mathcal{M}$ be the family of Monotone Boolean functions. Then there exists a $g \notin \mathcal{M}$ such that $\mathcal{M} \cup \{g\}$ is also* s-extremal.

Proof. Here is the function $g : \{0,1\}^n \to \{0,1\}$. Consider a string $w \in \{0,1\}^n$. We equivalently denote the function g as g_w because the function is parameterized by w.

$$g_w(x) = \begin{cases} 1, & x = w \text{ and } wt(w) = n-1 \\ 0, & otherwise \end{cases}$$

The alternation of the function g_w is 2 and hence not in the family. Upon adding this function, the family will additionally shatter one more set $S = \{w, 1^n\}, w \in \{0,1\}^n$. Earlier, it was family of monotone Boolean function, it could only obtain $\{\phi, \{1^n\}, \{w, 1^n\}\}$ but with the addition of g_w all the 4 subsets can be obtained. Observe that due to the peculiarity of the function g_w which we added, no other sets get shattered. Hence the family shatters exactly one extra set upon this addition. Also, observe that we can repeat the process of adding a function $n-1$ times in a similar way thus shattering a new set each time. □

4.5 Subspace VC-dimension

In this section, we study a more restricted version of VC-dimension which we call $Subspace$ VC-dimension or SVC where we shatter a subspace instead of a set.

From the definition of Subspace VC-dimension (Definition 2.1.6), we have the following observation.

Observation 4.5.1. *Let $\mathcal{F}$ be a family of Boolean functions. Then* $\mathsf{SVC}(\mathcal{F}) \leq \mathsf{VC}(\mathcal{F})$.

Proof. Let $\mathsf{SVC}(\mathcal{F}) = d$ and S be the shattered subspace. We can observe that the basis vectors of the subspace S is shattered by $\mathcal{F}$. Let T be the set shattered by $\mathcal{F}$. The proof follows from the fact that each basis vector of subspace in S corresponds to a subset of T. □

Similar to VC-dimension , one may want to see if Subspace VC-dimension bounds lead to some sample complexity bounds for a learning algorithm. We remark that there is no immediate connection with sample complexity bounds. We gave a brief sketch of Theorem 3.2.1 and can observe that the argument breaks down if we insist the shattered set to be a subspace.

We now show bounds on SVC for some well known families. Let $\mathcal{M}$ denote the family of monotone Boolean functions. We show,

Proposition 4.5.2. $\mathsf{SVC}(\mathcal{M}) = 0$

Proof. It suffices to argue that $\mathsf{SVC}(\mathcal{M}) < 1$. Suppose we can shatter a subspace $S \subseteq \{0,1\}^n$ where dimension of S is at least 1. Without loss of generality, we can assume that $S = \{0^n, w_1\}$ which is of dimension 1. Consider the subspace $\{0^n\}$. If we require $f \cap S = \{0^n\}$, then we need $f(0^n) = 1$. But since f is a monotone function, we also have $f(w_1) = 1$. Thus, $\mathsf{SVC}(\mathcal{M}) < 1$. □

Define the class

$$\mathcal{F} = \left\{ \neg \bigoplus_{i \in S} x_i \mid S \subseteq [n] \right\}$$

We call this class as family of parity.

Proposition 4.5.3. $\mathsf{SVC}(\mathcal{F}) = 2$

Proof. We will show the lower bound by shattering a subspace of dimension 2. Let $S = \{0^n + 0^{n-2}.w \mid w \in \{0,1\}^2\}$. We obtain all the subspaces of S from the function described below. We denote the right most bit of the n-bit string x by x_1 and left most by x_n.

Subspace of S	**Function**
$\{0^n, 0^{n-2}.01\}$	$\neg(x_2 \oplus x_3)$
$\{0^n, 0^{n-2}.10\}$	$\neg(x_1 \oplus x_3)$
$\{0^n, 0^{n-2}.11\}$	$\neg(x_1 \oplus x_2)$

We will use the following relation to prove the upper bound. Let U, V be two subspaces then the following relation holds.

$$dim(U \cap V) = dim(U) + dim(V) - dim(U + V) \tag{4.1}$$

where $dim(U)$-dimension of U. We refer the readers to the Linear Algebra textbook by Axler (1997) for details of the proof.

Let us suppose, U be the subspace of dimension d shattered by the family. Consider a function f in the family and a set $V \subseteq \{0,1\}^n$ such that $\forall x \in \{0,1\}^n,\ f(x) = 1 \Leftrightarrow x \in V$.

We also observe that V forms a subspace and has dimension $dim(V) = n - 1$ because f is of the form $\neg \bigoplus_{i \in S} x_i$ which can further be expressed in the form of a equation

$$\bigoplus_{i \in S} x_i = 0.$$

Since, f is set of just 1 linear equation, the dimension of V is $n - 1$.

To show that U is shattered, we need for all subspaces $U' \leq U$, $\exists f \in \mathcal{F}$ such that $f^{-1} \cap U = U'$. Here on, we denote f^{-1} by V. Consider the subspace $U \cap V$.

Using equation 4.1, we have,

$$dim(U \cap V) = d + n - 1 - dim(U + V)$$

$$\Rightarrow d - 1 \leq dim(U \cap V) \leq n \qquad (0 \leq dim(U + V) \leq n)$$

The above inequality says that for any $V \in \mathcal{F}$, $dim(U \cap V) \geq d - 1$. Thus we will not be able to shatter the subspace U because $\nexists V \in \mathcal{F}$ from which we can obtain subspaces of U smaller than $d - 1$. Hence, we conclude that $d \leq 2$.

□

We see Subspace VC-dimension of symmetric functions now.

Proposition 4.5.4. $\mathsf{SVC}(\textsc{Sym}) = \lfloor \log(n + 1) \rfloor$.

Before proving this proposition, we will first look at the following claim below.

Claim 4.5.5. *Any subspace which can be shattered cannot have two vectors of the same Hamming weight. This also implies that, the size of the largest subspace which can be shattered is at most* $n+1$.

Proof. Consider a subspace S which contains two elements w_1 and w_2 such that $wt(w_1) = wt(w_2)$. Consider the subspace $S_1 = \{0^n, w_1\}$. Any symmetric function which evaluates to 1 on w_1, will also evaluate to 1 on w_2. Thus, there does not exists a symmetric function g such that $g^{-1} \cap S = S_1$. Hence, the subspace S cannot be subspace-shattered. Thus, the largest shattered-subspace cannot have two vectors with same weight.

□

Proof for Proposition 4.5.4. *Lower Bound.* We will show the lower bound by giving the following subspace of dimension of $\lfloor \log(n+1) \rfloor$. Consider a set $S \subseteq \{0,1\}^n$ which has the following properties.

- $0^n \in S$
- $wt(x) = 2^k$, $k \in [\lfloor \log n \rfloor]$
- $\forall x, y \in S, x \wedge y = 0^n$

First, we want to remark that the set described above does exist. We include an element x in S such that it follows the first two property and $\forall y \in S$, $x \wedge y = 0^n$.

Clearly, we can see that every vectors in the set S has different weights. Hence for any $S' \leq S$, $\exists f \in \mathcal{F}$ such that $f \cap S = S'$ and this f is given by OR of symmetric functions of all the different weights in S'.

Upper Bound. From Claim 4.5.5, we know that largest possible set which can be shattered cannot have two vectors of same weight. Hence the subspace can have at most $n+1$ vectors. Let d be the dimension of this largest subspace. Hence $2^d \leq (n+1)$, which gives $d \leq \lfloor \log(n+1) \rfloor$. □

We now mention a motivation to further study the Subspace VC-dimension of Boolean function families.

Recall the definition of CC^0 and AC^0 complexity classes in Chapter 2. It is widely conjectured that the classes AC^0 and CC^0 are separate but have been a long-standing open problem. To show that the two classes of functions are different, consider the families in each class. If we succeed to show that their subspace VC-dimension is different as families, then that would separate these classes.

CHAPTER 5

VC Dimension of a Boolean Function

In this Chapter, we study VC-dimension of a Boolean function (family interpreted as set of characteristic vectors for the positive inputs of the function) defined in Definition 2.1.2 in Preliminaries (see Chapter 2). We prove VC-dimension bounds for several Boolean functions including *Monotone, k-slice, and Symmetric*. We show connection between $\mathsf{VC}(f)$ and *Certificate Complexity* of a Boolean function thus establishing relation with other complexity measures of Boolean functions such as *degree, sensitivity, Influence* etc. Note that our universe $U = [n]$ and family $\mathcal{F} \subseteq 2^U$.

5.1 VC-dimension Bounds of Specific Boolean Functions

We begin with the following upper bound of VC-dimension of any Boolean function.

Proposition 5.1.1. *For any Boolean function f, $\mathsf{VC}(f) \leq n$.*

Proof. This follows from the general known property of the VC-dimension that $\mathsf{VC}(\mathcal{F}) \leq \log(|\mathcal{F}|)$. Indeed, the latter is because, if we have to shatter a set S of size k, we need to have at least 2^k (the number of subsets of S) many sets in $\mathcal{F}$. Hence, a family $\mathcal{F}$ cannot shatter a set of size more than $\log(|\mathcal{F}|)$ and it implies that VC-dimension$(\mathcal{F}) \leq \log(|\mathcal{F}|)$. In the above setting, $|\mathcal{F}| \leq 2^n$ and the bound follows. □

Are there Boolean functions where the VC-dimension(f) is equal to n? Yes, the trivial function f where it is constant 1 always. Are there functions where $\mathsf{VC}(f) = 0$? Yes, Boolean functions where there is at most one input $x \in \{0,1\}^n$ where $f(x) = 1$ have $\mathsf{VC}(f) = 0$.

We first show VC-dimension bound of THRESHOLD function denoted by $\textsc{Th}_n^k$. Recall that, $\textsc{Th}_n^k$ function evaluates to 1 if and only if weight of the input is at least k.

Proposition 5.1.2. $\mathsf{VC}(\textsc{Th}_n^k) = n - k$.

Proof. We will show the lower bound by describing a set S of size $n-k$ which is shattered by $\mathcal{F}$. Consider $S = \{1, 2, \dots n-k\}$. Let S' be any subset of S. We have to find F such that $S \cap F = S'$. Define :

$$F = S' \cup W, \text{ where } W = \{n-k+1, n-k+2, \dots n\}$$

Observe that $|F| \geq k$ and hence, $F \in \mathcal{F}$. By construction, $W \cap S = \phi$ and hence, $F \cap S = S'$. Observe that any set of size $n-k$ can be shattered by the family TH_n^k. Now we will show $\mathsf{VC}(\text{TH}_n^k) \leq n-k$. We need to show that none of the sets S of size more than $n-k$ can be shattered. Let S be such a set. We show that S cannot be shattered by exhibiting a subset S' that cannot be obtained as $S \cap F$ for any $F \in \mathcal{F}$. We now prove this by contradiction. Suppose size of S is $n-k+1$ and $S' = \phi$. Let $F \in \mathcal{F}$ such that $S \cap F = S'$. Since $S' = \phi$, so F has no elements from S. Hence, $|F| \leq k-1$ which contradicts the fact that $F \in \mathcal{F}$ as for all $F \in \mathcal{F}$, $|F| \geq k$. □

We have the following immediate corollary about AND, OR and MAJORITY. AND is exactly TH_n^n, OR is TH_n^1 and MAJORITY is $\text{TH}_n^{\lceil n/2 \rceil}$.

Corollary 5.1.3. $\mathsf{VC}(\vee_n) = n-1$, $\mathsf{VC}(\wedge_n) = 0$ *and* $\mathsf{VC}(\text{MAJORITY}) = \lfloor n/2 \rfloor$.

We see VC-dimension bound for PARITY function now.

Proposition 5.1.4. $\mathsf{VC}(\oplus_n) = n-1$

Proof. We will first see the upper bound $\mathsf{VC}(\oplus_n) \leq n-1$. The $PARITY$ function is $TRUE$ on 2^{n-1} assignments. Hence, the size of the concept class will be 2^{n-1}. Hence, from Proposition 3.3.2 , we conclude that $\mathsf{VC}(\oplus_n) \leq n-1$. We now show the lower bound $\mathsf{VC}(\oplus_n) \geq n-1$ by showing that the family shatters a set $S \subseteq U$ of size $n-1$. Without loss of generality, consider $S = \{1, 2, \dots, n-1\}$. Let $S' \subseteq S$ be any arbitrary subset of S. We now find $F \in \mathcal{F}$ such that $S \cap F = S'$. Consider these 2 cases when size of S' is odd or even.

Case 1 : $|S'|$ is odd. In this case $F = S'$ because $|S'|$ is odd and hence, $S' \in \mathcal{F}$. Hence, $F \cap S = S'$.

Case 2 : $|S'|$ is even. In this case we have $F = S' \cup \{n\}$. $|S'|$ is even but F is odd and hence, $F \in \mathcal{F}$. Hence, $F \cap S = S'$.

□

We now give a trivial VC-dimension upper bound of a symmetric function.

Proposition 5.1.5. *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a symmetric Boolean function that evaluates to* 1 *on a set of weights $S = \{w_1, w_2, \ldots, w_k\}$. Then* $\mathsf{VC}(f) \leq max_{1 \leq i \leq k}\{w_i\}$.

Proof. Observe that the family cannot shatter a set of size greater than w_{max} because all the sets in the family is of size at most w_{max}. Hence, $\mathsf{VC}(f) \leq w_{max}$. □

We now compute VC-dimension of a monotone function which is related to *prime-implicant* of the function. An *implicant* of a function is a conjunction of literals[1] such that if the implicant is 1 then the function is 1. A *prime-implicant* is an *implicant* which is not absorbed by another *implicant* i.e. removal of any literal from the *prime implicant* should not give another *implicant*. Largest *prime-implicant* is the minimum number of literals which upon setting to 1, the function evaluates to 1.

Theorem 5.1.6. *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a monotone function. Let k be the size of a largest prime implicant of the function f then* $\mathsf{VC}(f) = n - k$.

Proof. Lower Bound: We prove the lower bound by shattering a set $S \subseteq U$ of size $n - k$. Suppose the largest prime implicant of f is k. A Boolean function f is monotone if and only if there is a monotone circuit[2] computing the function. Thus, the minimal DNF-form of the function f has only positive literals. Hence, the prime implicant is of the form $(x_{i_1} \wedge \ldots \wedge x_{i_k})$. Clearly the set $\{i_1, \ldots, i_k\} \in \mathcal{F}$ where $i_j \in [n]$. The set which will be shattered is, $S = U \setminus \{i_1, \ldots, i_k\}$. To obtain any subset $S' \subseteq S$, we give F such that $F \cap S = S'$ is as follows:-

$$F = S' \bigcup \{i_1, \ldots, i_k\}$$

It can be observed that $F \in \mathcal{F}$ as the function f is a monotone Boolean function. So all the supersets of the set $\{i_1, \ldots, i_k\}$ will be in the family because prime implicant is the minimum number of literals which upon setting to 1 makes the function evaluate to 1. Hence, we shattered a set S of size $n - k$.

[1] A literal is a variable in positive (x_i) or negative form ($\overline{x_i}$)

[2] A circuit with only $\wedge$ and $\vee$ gates.

Upper Bound : We show the upper bound by showing that no set of size greater than $n-k$ can be shattered. Suppose, the set $S \subseteq U$ be of size $n-k+1$. To obtain the empty set from the shattering of the set S, we need some $F \subseteq U \setminus S$. Clearly $|F| \leq n-(n-k+1) = k-1$. But we know all the sets in the family $\mathcal{F}$ has size at least k because largest prime implicant is of size k. Hence, no set of size greater than $n-k$ can be shattered. □

Since *k-slice* functions are also *monotone* Boolean function, and largest implicant is of size k, we have the following immediate corollary from Theorem 5.1.6.

Corollary 5.1.7. $\mathsf{VC}(k\text{-}slice) = n-k$.

5.2 Shattering-Extremal Boolean functions

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $\mathcal{F}$ be the corresponding family as defined in Chapter 2. We look at *s-extremal* family in this setting. Here is an explicit *s-extremal* family.

Proposition 5.2.1. *Let* TH_n^k *be a Threshold Boolean function.* TH_n^k *is s-extremal.*

Proof. For a Threshold Boolean function, $|f^{-1}(1)| = \sum_{i=k}^{n} \binom{n}{i} = \sum_{i=0}^{n-k} \binom{n}{i}$. The set of shattered sets by TH_n^k are all sets of size $n-k$ as seen in Proposition 5.1.2. Hence, $|Sh(f)| = \sum_{i=0}^{n-k} \binom{n}{i}$ which gives $|Sh(f)| = |f^{-1}(1)|$. □

The above proposition generalizes a number of *s-extremal* family such as *AND, OR, MAJ* etc. Here is another example of *s-extremal* family.

Proposition 5.2.2. *Let* $f : \{0,1\}^n \to \{0,1\}$ *be a negation of a monotone Boolean function. The family* $\mathcal{F}$ *corresponding to* f *is* s-extremal.

Proof. The family $\mathcal{F}$ corresponding to f is a subset closed family. Thus, for any set $S \in \mathcal{F}, \forall S' \subseteq S, \exists F \in \mathcal{F}$ such that $F \cap S = S'$. Hence all the sets present in the family can be shattered. Also observe that, any set G which is not present in the family $\mathcal{F}$ cannot be shattered because $\nexists F \in \mathcal{F}$ such that $F \cap G = G$. Hence the number of shattered sets is exactly equal to the size of the family. □

5.2.1 Strong Shattering Complexity of Boolean Functions

Recall the definition of 1-certificate from section 2.2. Interpreting strong shattering in the context of functions, we say that $S \subseteq [n]$ is *strongly shattered* if there is a 1-certificate of f which does not set any variable in S. Thus, if the function strongly-shatter a set of size d, then there is a 1-certificate of size $n - d$. Conversely, if there is a 1-certificate of size at most $n - d$, then we can strongly shatter a set of size d. This gives the following relation:

Proposition 5.2.3. *For any Boolean function f, the 1-certificate complexity (denoted by $C^1(f)$) and the strong shattering dimension (say $\mathsf{VC}_s(f)$) satisfy:*

$$C^1(f) + \mathsf{VC}_s(f) \geq n$$

Using $C(f) \geq C^1(f)$ and $\mathsf{VC}(f) \geq \mathsf{VC}_s(f)$ we have,

Proposition 5.2.4. *For any Boolean function f, the certificate complexity (denoted by $C(f)$) and the $\mathsf{VC}(f)$:*

$$C(f) + \mathsf{VC}(f) \geq n$$

The above Proposition 5.2.4 establishes a connection between VC-dimension and certificate complexity of a function, thus establishing a connection with all other measures of a Boolean function described in Table 3.1. We now show another application to give a lower bound on the degree of a monotone function. From Proposition 5.2.4 and Theorem 5.1.6 , we can derive that for a monotone Boolean function f, $C(f) \geq k$ where k is the largest prime implicant . Using this fact and $C(f) \leq \mathsf{deg}(f)^3$ by Midrijanis (2004), we have the following immediate corollary.

Corollary 5.2.5. *For any monotone function $f : \{0,1\}^n \rightarrow \{0,1\}$, we have $\mathsf{deg}(f) \geq k^{\frac{1}{3}}$ where k is the size of largest prime-implicant.*

5.2.2 Extremal Family with VC-dimension at most 1

Let $\mathcal{F}$ be an *s-extremal* family. Let $G_{\mathcal{F}}$ be the *inclusion* graph corresponding to a family $\mathcal{F}$. Recall that, an *inclusion graph* $G_{\mathcal{F}}$ for a family $\mathcal{F}$ is defined as a graph whose vertices are all the sets in $\mathcal{F}$ and there is a directed edge from S to T ($S \neq T$), labelled with

$j \in [n]$ when $T = S \cup \{j\}$. The *inclusion* graph is indeed a subgraph of the Boolean Hypercube.The following characterization is known.

Theorem 5.2.6 (Mészáros and Rónyai (2013))**.** *A family $\mathcal{F}$ is s-extremal and of* VC-*dimension at most 1 if and only if $G_{\mathcal{F}}$ is a tree and all labels on the edges are distinct.*

We have the following interesting proposition for *average sensitivity* of such families by interpreting the family as a Boolean function.

Proposition 5.2.7. *Let $\mathcal{F}$ be an* s-extremal *family with* VC*-dimension at most 1. Then* average sensitivity *of such functions is at least $n(1 - \frac{2}{2^n})$.*

Proof. We have from Theorem 5.2.6 that graph $G_{\mathcal{F}}$ is a tree with distinct edge labels. We calculate the influence of i^{th} index in the following two cases.

Case 1 When i is one of the labelled edges in the graph $G_{\mathcal{F}}$, then $f(x) = f(x^{\oplus i})$. Thus, $Inf_i(f) = 1 - \frac{2}{2^n}$ because two endpoints of this edge is not sensitive at i^{th} index.

Case 2 When i is not one of the labelled edges of the graph $G_{\mathcal{F}}$, then $f(x) \neq f(x^{\oplus i})$. Thus, $Inf_i(f) = 1$.

Hence, average sensitivity (also known as total influence) is at least $n(1 - \frac{2}{2^n})$.

□

5.3 Complexity of Computing VC-dimension

In this section, we look at the complexity of computing VC-dimension of a function. In Schaefer's definition (Schafer (1996)), the problem is shown to be Σ_3^p-complete. *Matrix* VC*-dimension* i.e. family represented as incidence matrix, is shown to be LOGNP-complete by Papadimitriou and Yannakakis (1996). We define the decision version of the problem in the characteristic-vectors definition first.

Definition 5.3.1 (VCD). Given a circuit C and a positive integer k, decide if the VC-dimension of the function computed by the circuit is at least k or not.

Proposition 5.3.2. VCD $\in \Sigma_3^p$

Proof. Consider we have been given a circuit C and a number k. The computation is to find whether the VC-dimension is at least k or not. Here is the characterization, which shows that VCD is in Σ_3^p.

$$\begin{aligned}\mathsf{VC}(C) \geq k \Leftrightarrow & (\exists i_1, i_2, \ldots i_k \in U) \\ & (\forall x \in \{0,1\}^k)(\exists F \in \mathcal{F}) \\ & (\forall j \in \{1, \ldots k\})\,(x_F \wedge x_S[i_j] = x[j])\end{aligned}$$

□

Proposition 5.3.3. VCD *is* NP*-hard.*

Proof. Consider a formula ϕ on n variables. We reduce SAT to VCD. We show that, $\phi \in \mathsf{SAT} \iff \mathsf{VC}(\phi') \geq \frac{n}{\log n}$. The formula after reduction is $\phi' = \phi \wedge (x_{n+1} \vee \ldots \vee x_{2n})$. To see the correctness of the reduction, observe that whenever ϕ is satisfiable then there will be at least $2^n - 1$ satisfying assignments for ϕ'. Hence, from the lower bound argument (see Proposition 3.3.4), VC-dimension$(\phi') \geq \frac{n}{\log n}$. If ϕ is not satisfiable then VC-dimension of the function given by ϕ' will be 0. Hence, VCD is NP-hard. □

Definition 5.3.4 (monotone-VCD). Given a circuit $\mathcal{C}$ which has only $\wedge$ and $\vee$ gates and a positive integer k, decide if the VC-dimension of the function computed by the circuit is at least k or not.

Theorem 5.3.5. $monotone$-VCD *is* NP*-complete.*

Proof. From Theorem 5.1.6, it can be concluded that problem of finding VC-dimension of a monotone function f is equivalent to finding the largest *prime-implicant* of the function f. Given a problem and a certificate set of k-variables, one can easily verify whether the size of *prime-implicant* is k or not by setting those variables to 1 and checking whether the function evaluates to 1 or not. Hence, the problem is in NP.

Manquinho *et al.* (1997) proves that the problem of finding *prime-implicant* size of a monotone Boolean function is NP-hard. The result also appears in the textbook by Nguyen (2005). Thus, $monotone$-VCD is NP-complete. □

We were able to show that VC-dimension computation for a single Boolean function is NP-hard but we gave a characterization to put the problem in Σ_3^p. We believe it to be hard for the class Σ_3^p. But, on the promise that the function is monotone, we were able to show that the VC-dimension computation is NP-complete using the VC-dimension bound of a monotone function.

CHAPTER 6

Conclusion and Future Work

In this thesis, we studied several variants of VC-dimension , firstly for families of Boolean functions in Chapter 4 and then for a single Boolean function in Chapter 5. Due to its applicability in sample complexity bounds for any learning algorithm, we computed VC-dimension of several families such as *k-slice, Read-once, family of functions with alternation* 1 and generalized to *alternation* k family. Non-monotone Boolean functions were a center of our main focus. We defined a family of non-monotone functions parameterized with maximum alternation and computed an asymptotically matching lower and upper bound of VC-dimension in Chapter 4. We gave an application of this bound by giving a $\Omega(dk)$ bound for a k-union family. The construction of k-union family is non-geometric and explicit. Family of k-union is a long-studied problem by Blumer *et al.* (1989), Reyzin (2006) and Eisenstat and Angluin (2007). We also showed *s-extremal* properties of *Monotone* family and verified Mészáros and Rónyai (2013) conjecture for the same.

We then defined a family corresponding to a Boolean function based on its characteristic vectors in Chapter 5 and computed VC-dimension of several functions such as *Monotone, k-slice, Symmetric* etc. We gave a characterization to put the computational problem of deciding the VC-dimension in Σ_3^p and showed that it was **NP**-hard to compute. There is no connection between VC-dimension and size or depth of the circuit computing a function f as there exist small circuits with high VC-dimension as well as low VC-dimension and vice-versa. But we showed its connection to a combinatorial complexity measure of Boolean function called certificate complexity in Chapter 5.

For future directions, we leave the following questions for each of these variants to ponder.

- We gave bounds for k-closure of families for $k \leq \Theta(\sqrt{n})$. A natural question is to show unconditional bound for any values of k.
- We gave a VC-dimension bound for k-fold union family to be $O(dk)$. Eisenstat and Angluin (2007) gave a probabilistic construction of k-fold union family matching the VC-dimension bound $O(dk \log k)$. Thus, an explicit and non-geometric bound matching $O(dk \log k)$ remains to prove for a k-fold union family.

- For *s-extremal* family, apart from the *Meszaros-Ronyai* Conjecture, it will be interesting to find more naturally occurring *s-extremal* families, and investigate if this property can be used to improve *Sauer-Shelah Lemma* bound, *bound on Boolean closure of k families* etc.

- On the learning front, Theorem 3.2.5 is a way to establish a bridge between VC-dimension and Boolean function measures but we could not use it due to the inherent behavior of the learning algorithms. As one can see from Theorem 3.2.1, the bound is applicable for any learning algorithms with arbitrary distribution $\mathcal{D}$ but Theorem 3.2.5 insists on using Uniform distribution. However, this definitely gives a new avenue to try and possibly show interesting connections.

- VC-dimension of Boolean function has found applications in many areas as described in Section 3.6. We gave one relation with certificate complexity. Its connection with other circuit complexity measures is yet to be explored.

- All the 2^{2^n} Boolean functions can be partitioned based on the family of sets they shatter. For each Boolean function, $Sh(f)$ which is the set of shattered sets, is subset closed. Thus, $Sh(f)$ can be interpreted as a negation of a monotone function and each partition corresponds to a negation of a monotone function. For all these partitions, it will be interesting to investigate the common traits shared by the functions apart from VC-dimension and Interpolation degree.

REFERENCES

1. **Amireddy, P.**, **S. Jayasurya**, and **J. Sarma**, Power of Decision Trees with Monotone Queries. *In* **D. Kim**, **R. N. Uma**, **Z. Cai**, and **D. H. Lee** (eds.), *Computing and Combinatorics - 26th International Conference, COCOON 2020, Atlanta, GA, USA, August 29-31, 2020, Proceedings*, volume 12273 of *Lecture Notes in Computer Science*. Springer, 2020.

2. **Arora, S.** and **B. Barak**, *Computational Complexity:A Modern Approach*. Cambridge University Press, 2009.

3. **Axler, S. J.**, *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, New York, 1997. ISBN 0387982582.

4. **Barnett, J.**, **H. Correia**, **P. Johnson**, **M. Laughlin**, and **K. Wilson** (2010). Darwin Meets Graph Theory on a Strange Planet: Counting Full n-ary Trees with Labeled Leaves. *Alabama Journal of Mathematics*, **35**.

5. **Ben-David, S.** (2015). Notes on Classes with Vapnik-Chervonenkis Dimension 1.

6. **Bhattacharya, A.**, **S. Chakraborty**, **A. Ghosh**, **G. Mishra**, and **M. Paraashar**, Disjointness Through the Lens of Vapnik–Chervonenkis Dimension: Sparsity and Beyond. *In* **J. Byrka** and **R. Meka** (eds.), *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*, volume 176 of *Leibniz International Proceedings in Informatics (LIPIcs)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020. ISBN 978-3-95977-164-1. ISSN 1868-8969.

7. **Blais, E.**, **C. L. Canonne**, **I. C. Oliveira**, **R. A. Servedio**, and **L.-Y. Tan**, Learning Circuits with few Negations. *In* **N. Garg**, **K. Jansen**, **A. Rao**, and **J. D. P. Rolim** (eds.), *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2015. ISBN 978-3-939897-89-7. ISSN 1868-8969.

8. **Blumer, A.**, **A. Ehrenfeucht**, **D. Haussler**, and **M. K. Warmuth** (1989). Learnability and the Vapnik-Chervonenkis Dimension. *J. ACM*, **36**(4), 929–965. ISSN 0004-5411.

9. **Bollobás, B.** and **A. Radcliffe** (1995). Defect Sauer Results. *Journal of Combinatorial Theory, Series A*, **72**(2), 189–208. ISSN 0097-3165.

10. **Cook, S.**, **C. Dwork**, and **R. Reischuk** (1986). Upper and Lower Time Bounds for Parallel Random Access Machines without Simultaneous Writes. *SIAM Journal on Computing*, **15**(1), 87–97.

11. **Crama, Y.** and **P. L. Hammer**, *Boolean Functions : Theory, Algorithms and Applications*. 2011. ISBN 9780521847513.

12. **Dinesh, K.** and **J. Sarma** (2018). Alternation, Sparsity and Sensitivity: Bounds and Exponential Gaps . *Theoretical Computer Science*. ISSN 0304-3975.

13. **Dvir, Z.** and **S. Moran** (2018). A sauer-shelah-perles lemma for sumsets. *ArXiv*, **abs/1806.05737**.

14. **Ehrenfeucht, A.**, **D. Haussler**, **M. Kearns**, and **L. Valiant** (1989). A General Lower Bound on the Number of Examples Needed for learning. *Information and Computation*, **82**(3), 247 – 261. ISSN 0890-5401.

15. **Eisenstat, D.** and **D. Angluin** (2007). The VC Dimension of K-Fold Union. *Inf. Process. Lett.*, **101**(5), 181–184. ISSN 0020-0190.

16. **Gurvits, L.**, Linear Algebraic proofs of VC-Dimension based inequalities. *In* **S. Ben-David** (ed.), *Computational Learning Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997. ISBN 978-3-540-68431-2.

17. **Haussler, D.** and **E. Welzl** (1987). ϵ-nets and Simplex Range Queries. *Discrete & Computational Geometry*, **2**(2), 127–151.

18. **Huang, H.** (2019). Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, **190**(3), 949–955. ISSN 0003486X, 19398980.

19. **Jukna, S.** and **A. Lingas**, Lower Bounds for DeMorgan Circuits of Bounded Negation Width. *In* **R. Niedermeier** and **C. Paul** (eds.), *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*, volume 126 of *Leibniz International Proceedings in Informatics (LIPIcs)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019. ISBN 978-3-95977-100-9. ISSN 1868-8969.

20. **Klauck, H.**, On Quantum and Probabilistic Communication: Las Vegas and One-Way Protocols. *In Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, STOC '00. Association for Computing Machinery, New York, NY, USA, 2000. ISBN 1581131844.

21. **Kleitman, D.** and **G. Markowsky** (1975). On Dedekind's Problem: The Number of Isotone Boolean Functions. . *Transactions of the American Mathematical Society*, **213**, 373–390. ISSN 00029947.

22. **Koiran, P.**, VC dimension in circuit complexity. *In Proceedings of Computational Complexity (Formerly Structure in Complexity Theory)*. 1996.

23. **Koroth, S.** and **J. Sarma** (2017). Depth lower bounds against circuits with sparse orientation. *Fundam. Informaticae*, **152**(2), 123–144.

24. **Kozen, D. C.**, *Theory of Computation*. Springer-Verlag London, 2006.

25. **Kremer, I.**, **N. Nisan**, and **D. Ron**, On Randomized One-Round Communication Complexity. *In Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95. Association for Computing Machinery, New York, NY, USA, 1995. ISBN 0897917189.

26. **Kushilevitz, E.** and **N. Nisan**, *Communication Complexity*. Cambridge University Press, 1996.

27. **Lin, C.** and **S. Zhang**, Sensitivity Conjecture and Log-Rank Conjecture for Functions with Small Alternating Numbers. *In ICALP*. 2017.

28. **Linial, N.**, **Y. Mansour**, and **R. L. Rivest** (1991). Results on learnability and the Vapnik-Chervonenkis dimension. *Information and Computation*, **90**(1), 33–49. ISSN 0890-5401.

29. **Manquinho, V.**, **P. Flores**, **J. Silva**, and **A. Oliveira**, Prime Implicant Computation using Satisfiability Algorithms. 1997. ISBN 0-8186-8203-5.

30. **Markov, A. A.** (1958). On the Inversion Complexity of a System of Functions. *Journal of ACM*, **5**(4), 331–334. ISSN 0004-5411.

31. **Mészáros, T.** and **L. Rónyai** (2013). Shattering-extremal Set Systems of Small VC-dimension. *International Scholarly Research Notices*, **2013**.

32. **Midrijanis, G.** (2004). Exact Quantum Query Complexity for total Boolean Functions. *arXiv: Quantum Physics*.

33. **Mixon, D. G.** and **J. Peterson** (2015). Learning Boolean Functions with Concentrated Spectra. *CoRR*, **abs/1507.04319**.

34. **Mossel, E.** and **C. Umans** (2002). On the Complexity of Approximating the VC Dimension. *Journal of Computer and System Sciences*, **65**(4), 660–671. ISSN 0022-0000. Special Issue on Complexity 2001.

35. **Mészáros, T.** and **L. Rónyai** (2014). Shattering-Extremal Set Systems of VC Dimension at most 2. *Electronic Journal of Combinatorics*, **21**.

36. **Natschläger, T.** and **M. Schmitt** (1996). Exact VC-dimension of Boolean monomials. *Information Processing Letters*, **59**(1), 19 – 20.

37. **Nguyen, H. S.**, Approximate Boolean Reasoning Approach to Rough Sets and Data Mining. *In* **D. Ślęzak**, **J. Yao**, **J. F. Peters**, **W. Ziarko**, and **X. Hu** (eds.), *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-31824-8.

38. **Nisan, N.** (1991). CREW PRAMs and Decision Trees. *SIAM Journal on Computing*, **20**(6), 999–1007.

39. **Nisan, N.** and **M. Szegedy**, On the Degree of Boolean Functions as Real Polynomials. *In Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92. Association for Computing Machinery, New York, NY, USA, 1992. ISBN 0897915119.

40. **Pajor, A.**, *Sous-espaces LN/L Des Espaces de Banach / Alain Pajor* . Hermann Paris, 1985. ISBN 2705660216.

41. **Papadimitriou, C. H.** and **M. Yannakakis** (1996). On Limited Nondeterminism and the Complexity of the V-C Dimension. *Journal of Computer and System Sciences*, **53**(2), 161–170. ISSN 0022-0000.

42. **Procaccia, A. D.** and **J. Rosenschein** (2006). Exact VC-Dimension of Monotone Formulas. *Neural Information Processing -Letters and Reviews*, **10**.

43. **Rao, A.** and **A. Yehudayoff**, *Communication Complexity: and Applications*. Cambridge University Press, 2020.

44. **Reyzin, L.**, Lower Bounds on the VC Dimension of Unions of Concept Classes. 2006.

45. **Rivest, R. L.** (1987). Learning Decision Lists. *Machine Learning*, **2**(3), 229–246.

46. **Roy, A.** and **J. Sarma**, On Alternation, VC-dimension and k-fold Union of Sets. *In* **J. Nešetřil**, **G. Perarnau**, **J. Rué**, and **O. Serra** (eds.), *Extended Abstracts EuroComb 2021*. Springer International Publishing, Cham, 2021. ISBN 978-3-030-83823-2.

47. **Sauer, N.** (1972). On the Density of Families of Sets. *Journal of Combinatorial Theory, Series A*, **13**(1), 145–147. ISSN 0097-3165.

48. **Schafer, M.**, Deciding the Vapnik-Cervonenkis dimension is Sigmaˆ3-complete. *In Proceedings of Computational Complexity (Formerly Structure in Complexity Theory)*. 1996.

49. **Shelah, S.** (1972). A Combinatorial Problem; Stability and Order for Models and Theories in Infinitary Languages. *Pacific Journal of Mathematics*, **41**(1), 247 – 261.

50. **Shpilka, A.** and **A. Yehudayoff** (2010). Arithmetic circuits: A survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.*, **5**(3–4), 207–388. ISSN 1551-305X.

51. **Sontag, E.**, VC Dimension of Neural Networks. *In* **C. Bishop** (ed.), *Neural Networks and Machine Learning*. Springer, Berlin, 1998, 69–95.

52. **Spencer, J.** and **L. Florescu**, *Asymptopia*. Student mathematical library. American Mathematical Society, 2014. ISBN 9781470438432.

53. **Tal, A.**, Properties and Applications of Boolean Function Composition. *In Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13. Association for Computing Machinery, New York, NY, USA, 2013. ISBN 9781450318594.

54. **Valiant, L. G.** (1984). A Theory of the Learnable. *Commun. ACM*, **27**(11), 1134–1142.

55. **Vapnik, V. N.** and **A. Y. Chervonenkis** (1971). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, **16**(2), 264–280.

56. **Yao, A. C.-C.**, Some Complexity Questions Related to Distributive Computing(Preliminary Report). *In Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79. Association for Computing Machinery, New York, NY, USA, 1979. ISBN 9781450374385.

CURRICULUM VITAE

1. **NAME** : Amit Roy

2. **DATE OF BIRTH** : 17 March 1997

3. **EDUCATIONAL QUALIFICATIONS**

2018 **Bachelor of Technology (B.Tech.)**

Institution : Institute of Engineering and Management, Kolkata

Specialization : Computer Science and Engineering

2021 **Master of Science (M. S.)**

Institution : Indian Institute of Technology, Madras

Specialization : Computer Science and Engineering

GENERAL TEST COMMITTEE

CHAIRPERSON	:	Dr. Balaraman Ravindran Professor Department of Computer Science, Indian Institute of Technology, Madras
GUIDE(S)	:	Dr. Jayalal Sarma Professor Department of Computer Science, Indian Institute of Technology, Madras
MEMBERS	:	Dr. Yadu Vasudev Assistant Professor Department of Computer Science, Indian Institute of Technology, Madras
		Dr. Jayanthan A V Professor Department of Mathematics, Indian Institute of Technology, Madras