

Some new maximum VC classes

Hunter R. Johnson^{*†‡}

September 8, 2013

Abstract

Set systems of finite VC dimension are frequently used in applications relating to machine learning theory and statistics. Two simple types of VC classes which have been widely studied are the maximum classes (those which are extremal with respect to Sauer’s lemma) and so-called Dudley classes, which arise as sets of positivity for linearly parameterized functions. These two types of VC class were related by Floyd, who gave sufficient conditions for when a Dudley class is maximum. It is widely known that Floyd’s condition applies to positive Euclidean halfspaces and certain other classes, such as sets of positivity for univariate polynomials.

In this paper we show that Floyd’s lemma applies to a wider class of linearly parameterized functions than has been formally recognized to date. In particular we show that, modulo some minor technicalities, the sets of positivity for any linear combination of real analytic functions is maximum on points in general position. This includes sets of positivity for multivariate polynomials as a special case.

1 Introduction

Maximum set systems are in some sense the most perfect systems of finite VC dimension. They arise most notably from the systems given by “positive” half-spaces in Euclidean space. They also arise as the dual set system associated with a simple arrangement of hyperplanes. Their desirable features include a certain kind of recursive structure which allows for, among other things, the existence of so-called sample compression schemes, and as such they are central to most approaches to proving the long outstanding sample compression conjecture [1, 6, 7, 15, 12]. Some further uses of maximum set systems exist in machine learning and model theory [13, 8, 9].

In what follows we first provide the definitions for the basic notions of interest, including set systems, VC dimension, the maximum property and linearly parameterized set systems. We then go on to establish our results in the subsequent section.

Our results relate to two criteria given by Floyd which are sufficient for a linearly parameterized set system to have the maximum property. While several specific applications of Floyd’s theorem have been given, other powerful applications seem to have been overlooked. In particular, there seems to be no mention in the literature that Floyd’s result applies to general multivariate (rather than univariate) polynomials. More generally we note the important fact that any linear combination of analytic functions satisfies Floyd’s criteria.

2 Basic definitions

Let X be a set and $\mathcal{P}(X)$ its power set. We call any $\mathcal{C} \subseteq \mathcal{P}(X)$ a *set system* on X . For any $X_0 \subseteq X$, we let $\mathcal{C}|_{X_0}$ denote $\{C \cap X_0 : C \in \mathcal{C}\}$. We say that \mathcal{C} *shatters* $X_0 \subseteq X$ if $\mathcal{C}|_{X_0} = \mathcal{P}(X_0)$.

^{*}John Jay College, CUNY, 899 10th Ave New York, NY 10019

[†]hujohnson@jjay.cuny.edu

[‡]Keywords: combinatorial problems, computational geometry, VC dimension

The *Vapnik-Chervonenkis (VC) dimension* [20] of \mathcal{C} , when \mathcal{C} is non-empty, is defined as

$$\text{VC}(\mathcal{C}) = \sup\{|X_0| : X_0 \subseteq X \text{ is shattered by } \mathcal{C}\}.$$

When $\mathcal{C} = \emptyset$ we will use the convention that $\text{VC}(\mathcal{C}) = -1$.

If $\text{VC}(\mathcal{C})$ is finite then \mathcal{C} is said to be a *VC-class*. For natural numbers n and k , define

$$\binom{n}{\leq k} = \begin{cases} \sum_{i=0}^k \binom{n}{i} & \text{if } n \geq k \\ 2^k & \text{if } n < k. \end{cases}$$

A key combinatorial fact relating to VC classes is Sauer's lemma [16, 17].

Lemma 2.1. *Let $\alpha = \text{VC}(\mathcal{C})$. Then for any $X_0 \subseteq X$*

$$|\mathcal{C}|_{X_0} \leq \binom{|X_0|}{\leq \alpha}.$$

We say that \mathcal{C} is *maximum* [21] of VC dimension α if for any finite $X_0 \subseteq X$

$$|\mathcal{C}|_{X_0} = \binom{|X_0|}{\leq \alpha}.$$

Many set systems arise naturally as the family of sets defined by a parameterized formula in a mathematical structure. For instance, let X be a set, A a parameter set, and $f : X \times A \rightarrow \mathbb{R}$ a real-valued function. We use the notation $f_a : X \rightarrow \mathbb{R}$ to represent the function defined by $x \mapsto f(x, a)$. Let $\text{pos}(f_a) = \{x \in X : f_a(x) > 0\}$ and $\text{Pos}(f) = \{\text{pos}(f_a(x)) : a \in A\}$. Then $\text{Pos}(f)$ is a set system on X and has a well-defined VC dimension.

An interesting case occurs when f parameterizes a vector space of real-valued functions. Specifically, suppose that $f_i : X \rightarrow \mathbb{R}$ for $i = 1, 2, \dots, n$ are linearly independent real-valued functions, and $f_0 : X \rightarrow \mathbb{R}$ is a real-valued function. Let $\mathcal{F} = \{a_1 f_1(x) + a_2 f_2(x) + \dots + a_n f_n(x) : a_1, \dots, a_n \in \mathbb{R}\}$, and define $f_0(x) - \mathcal{F}$ to mean $\{f_0(x) - f(x) : f \in \mathcal{F}\}$. Then \mathcal{F} is a real vector space, and $f_0(x) - \mathcal{F}$ is an affine real vector space. We will use $\text{Pos}(f_0 - \mathcal{F})$ to denote $\{\text{pos}(f_0 - f) : f \in \mathcal{F}\}$. Set systems of the form $\text{Pos}(f_0 - \mathcal{F})$ have been called *Dudley classes* [1].

The following theorem is due to Dudley [22, 4]. Cover proved a similar (non-affine) result in [2].

Theorem 2.1. *If \mathcal{F} is an n -dimensional real vector space of real-valued functions defined on X , and $f_0 : X \rightarrow \mathbb{R}$, then $\text{VC}(\text{Pos}(f_0 - \mathcal{F})) = n$.*

Dudley classes include some natural set systems such as balls in Euclidean space, halfspaces in Euclidean space, and sets of positivity for polynomials, for which the coefficients are regarded as parameters. The following example is due to Dudley [3].

Example: We will show that balls in Euclidean 2-space (disks) form a Dudley class. The scheme of the example can be generalized to higher dimensions. Define $f_0(x, y) = -x^2 - y^2$ and $f(x, y) = a_3 y + a_2 x + a_1$. Then $f_0 - f \in f_0 - \mathcal{F}$ where $\mathcal{F} = \{a_3 y + a_2 x + a_1 : a_1, a_2, a_3 \in \mathbb{R}\}$. Note that $\text{pos}(f_0 - f)$ describes a disk with center $(\frac{-a_2}{2}, \frac{-a_3}{2})$ and radius $\sqrt{(\frac{a_2}{2})^2 + (\frac{a_3}{2})^2 - a_1}$.¹ Thus $\text{Pos}(f_0 - \mathcal{F})$ is the set system of all disks in the plane. Since it is also a 3-dimensional Dudley class, we can conclude from Theorem 2.1 that the VC dimension of the set of disks in the plane is 3. \square

The main link between maximum set systems and Dudley classes is due to Floyd [6] (Theorem 8.2).

¹When $(\frac{a_2}{2})^2 + (\frac{a_3}{2})^2 - a_1 < 0$, the radius does not exist; in this case $f_0 - f$ has no real solutions and $\text{pos}(f_0 - f)$ describes the empty-set. Including this degenerate case does not affect the VC dimension, because the empty-set can always be approximated by a sufficiently small disk.

Lemma 2.2. *Let \mathcal{F} be a vector space of real-valued functions on a set X , with $\dim(\mathcal{F}) = n$. Let $f_0(x)$ a function on X . Suppose further that*

1. *For any $A \subseteq X$ with $|A| = n$, the dimension of \mathcal{F} restricted to A is n .*
2. *For any $f \in \mathcal{F}$, there are at most n zeros of $f_0 - f$ in X .*

Then $\text{pos}(f_0 - \mathcal{F})$ is maximum of VC dimension n on X .

The assumption (1) above is (by Dudley's theorem) equivalent to the requirement that $\text{pos}(f_0 - \mathcal{F})$ shatter every set of size n in X , which is a necessary condition for the maximum property. This will be satisfied for an ordinary univariate polynomial $y = a_0 + a_1x + \dots + a_nx^n$ if X projects 1-1 onto the x -axis.

The assumption (2) above requires that $f_0 \notin \mathcal{F}$, since otherwise the space $f_0 - \mathcal{F}$ includes the constantly zero function. If assumption (2) is not respected, Euclidean halfspaces provide a counter-example to the lemma, as observed in [1].

We now proceed to give arguably more natural criteria which guarantee that a linear system of real-valued functions satisfy Floyd's conditions. For instance if f_0, \dots, f_n are linearly independent analytic functions and $X \subset \mathbb{R}^k$ is in general position, then with $\mathcal{F} = \text{span}\langle f_1, \dots, f_n \rangle$, Floyd's lemma applies to $f_0 - \mathcal{F}$.

This gives examples of maximum families which have not been given in the literature to date. Some examples are given at the end of the next section.

3 Results

In this section we will introduce topological and analytic conditions on \mathcal{F} and X which are sufficient to guarantee that $\text{Pos}(f_0 - \mathcal{F})$ is a maximum set system when restricted to subsets of X which are in general position (Theorem 3.1). The basic strategy for proving Theorem 3.1 is to associate subsets of X of size N with elements of X^N , and observe that the elements not satisfying Floyd's criteria constitute a thin part of X^N .

The elements of X^N on which Floyd's conditions fail will be seen to lie on the zero sets of certain functions arising as determinants of matrices. Establishing that these determinants do indeed have thin zero sets is the aim of Lemma 3.1.

When \mathcal{F} consists of analytic functions (which are defined before Proposition 3.1) the elements of X^N not satisfying Floyd's criteria will actually have Lebesgue measure zero. This implies that if a finite $X_0 \subseteq X$ is selected according to one of several common probability distributions, including the uniform and Gaussian distributions, then $\text{Pos}(f_0 - \mathcal{F})$ will almost surely be maximum when restricted to X_0 (see Corollary 3.3).

3.1

Let \mathcal{F} be an n -dimensional vector space of real-valued functions on a topological space X . We will say that \mathcal{F} is *admissible* if for any $f \in \mathcal{F}$,

1. f is continuous
2. If $f^{-1}(0)$ has non-empty interior, then f is constantly zero.

Note that any subspace of an admissible \mathcal{F} is admissible.

Equip $X^n = \underbrace{X \times X \times \dots \times X}_{n \text{ times}}$ with the product topology.

Lemma 3.1. Suppose \mathcal{F} is admissible and $f_1(x), \dots, f_n(x)$ is a basis for \mathcal{F} . Let $F : X^n \rightarrow \mathbb{R}$ be given by

$$F(x_1, \dots, x_n) = \det \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_n(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_n(x_n) \end{pmatrix}$$

then $F^{-1}(0) \subseteq X^n$ has empty interior.

Proof. The argument is by induction on n . If $n = 1$, the lemma follows from the assumptions on \mathcal{F} .

Suppose the lemma is known to hold for $n - 1$ and consider $F(x_1, \dots, x_n)$. Then

$$F = f_1(x_1) \begin{vmatrix} f_2(x_2) & \cdots & f_n(x_2) \\ \vdots & \ddots & \vdots \\ f_2(x_n) & \cdots & f_n(x_n) \end{vmatrix} + \cdots + f_n(x_1)(-1)^{1+n} \begin{vmatrix} f_1(x_2) & \cdots & f_{n-1}(x_2) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \cdots & f_{n-1}(x_n) \end{vmatrix}$$

where the vertical bars denote the determinant.

Suppose $U \subset F^{-1}(0)$ is open. Assume, by way of contradiction, that U is nonempty. Let V be the projection of U onto x_2, \dots, x_n . By inductive hypothesis, there is some $(a_2, \dots, a_n) \in V$ such that

$$\begin{vmatrix} f_2(a_2) & \cdots & f_n(a_2) \\ \vdots & \ddots & \vdots \\ f_2(a_n) & \cdots & f_n(a_n) \end{vmatrix} \neq 0$$

This gives

$$F(x_1, a_2, \dots, a_n) = c_1 f_1(x_1) + \cdots + c_n f_n(x_1)$$

for real numbers c_1, \dots, c_n , corresponding to the subdeterminants, and $c_1 \neq 0$. Define $U_{a_2, \dots, a_n} = \{a \in X : (a, a_2, \dots, a_n) \in U\}$. Then U_{a_2, \dots, a_n} is non-empty and open, and on this open set $F(x_1, a_2, \dots, a_n) = 0$. But $F(x_1, a_2, \dots, a_n) \in \mathcal{F}$, and therefore $F(x_1, a_2, \dots, a_n) = 0$ everywhere. This contradicts the linear independence of f_1, \dots, f_n , because $c_1 \neq 0$. Thus $F^{-1}(0)$ has empty interior. \square

Proposition 3.1, below, appears in [5] on p. 240. The statement given there is for the more general context of Banach spaces. The proof uses the technique of approximate differentiation; we will give a more elementary argument.

Recall that an infinitely differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *analytic* if for every x in the domain of f there is an open set U with $x \in U$ such that f is equal to its Taylor series expansion on U . We will use the fact that if $f : \mathbb{R} \rightarrow \mathbb{R}$ is analytic and takes non-zero values, then its zeros form a countable set [10].

Proposition 3.1. Let $f : \mathbb{R}^k \rightarrow \mathbb{R}$ be analytic. Suppose f is not constantly zero and let $A = f^{-1}(0)$. Then $\lambda(A) = 0$ where λ is Lebesgue measure.

Proof. The proof is by induction on k . Suppose $k = 1$. Then A is countable and therefore $\lambda(A) = 0$. If $k > 1$ then let χ_A be the indicator function for A .

Define $A_r = A \cap B_r$ where B_r is a k -ball of radius r centered at the origin. Then the functions $\{\chi_{A_n} : n \in \mathbb{N}\}$ converge monotonically to χ_A . By the monotone convergence theorem [14], $\lim_{n \rightarrow \infty} \int \chi_{A_n} = \int \chi_A$. Thus it suffices to show that $\int \chi_{A_n} = 0$ for all n . By replacing A with A_n if necessary, we may assume without loss that $\lambda(A) < \infty$.

Since χ_A takes only non-negative values, $\int_{\mathbb{R}^k} |\chi_A| d(x_1, \dots, x_k) = \int_{\mathbb{R}^k} \chi_A d(x_1, \dots, x_k) = \lambda(A) < \infty$. Thus χ_A satisfies the conditions of Fubini's theorem [5]. By Fubini's theorem we may evaluate $\int_{\mathbb{R}^k} \chi_A d(x_1, \dots, x_k)$ by iterated integration. From iterated integration and induction it is seen that $\int_{\mathbb{R}^k} \chi_A d(x_1, \dots, x_k) = 0$. This implies that $\lambda(A) = 0$. \square

Corollary 3.1. *If \mathcal{F} is a real vector space of real-valued functions defined on \mathbb{R}^k and \mathcal{F} has a basis consisting of real analytic functions then \mathcal{F} is admissible.*

Proof. Lebesgue measure zero implies empty interior. □

Corollary 3.2. *Suppose that \mathcal{F} is a real vector space of real-valued functions defined on \mathbb{R}^k for some $k \in \mathbb{N}$ and \mathcal{F} has a basis of real analytic functions. Then $F^{-1}(0)$ has Lebesgue measure zero, where F is as in Lemma 3.1.*

Proof. Note that F is analytic and not constantly zero, and therefore Proposition 3.1 applies. □

We will call a non-empty topological space X a *Baire space* if any countable union of closed sets with empty interior has empty interior. For natural numbers n and N , by $[N]^n$ we mean the subsets of $\{1, \dots, N\}$ of cardinality n . We will abuse notation slightly by writing $\langle i_1, \dots, i_n \rangle \in [N]^n$ to mean that $\{i_1, \dots, i_n\} \in [N]^n$ and $i_1 < \dots < i_n$. For an ordered set $q \in X^N$, we regard q as the function with domain $[N]$ and codomain X defined by $i \mapsto q_i$. Thus by $\text{range}(q)$ we mean the elements of X occurring in the ordered set q .

Theorem 3.1. *Suppose f_0, f_1, \dots, f_n are linearly independent real-valued functions defined on an infinite topological space X with the property that for every $N \in \mathbb{N}$, X^N is Baire in the product topology. Put $\mathcal{F} = \text{span}\langle f_0, f_1, \dots, f_n \rangle$ and $\mathcal{C} = \text{pos}(f_0 - \text{span}\langle f_1, \dots, f_n \rangle)$. Then if \mathcal{F} is admissible then for every $N > n$ there is $X_0 \subseteq X$ with $|X_0| = N$ such that $\mathcal{C}|_{X_0}$ is maximum of VC dimension n .*

Proof. Let $N \in \mathbb{N}$ be given. Let x_1, \dots, x_N be variables ranging over X . We can express the statement that x_1, \dots, x_N satisfy conditions (1) and (2) of Floyd's lemma using determinants. For any $B = \langle i_1, \dots, i_n \rangle \in [N]^n$, let F_B denote the function

$$F_B(x_1, \dots, x_N) = \det \begin{pmatrix} f_1(x_{i_1}) & f_2(x_{i_1}) & \cdots & f_n(x_{i_1}) \\ f_1(x_{i_2}) & f_2(x_{i_2}) & \cdots & f_n(x_{i_2}) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_{i_n}) & f_2(x_{i_n}) & \cdots & f_n(x_{i_n}) \end{pmatrix}.$$

Note that F_B ignores variables not in B .

Condition (1) will be true if for every $B \in [N]^n$, $F_B(x_1, \dots, x_N) \neq 0$. Note that for each choice of $B \in [N]^n$, $F_B^{-1}(0)$ has empty interior as a subset of X^N , as a consequence of Lemma 3.1.

Condition (2) of Floyd's lemma will be satisfied if the system given by

$$\begin{pmatrix} f_1(x_{i_1}) & f_2(x_{i_1}) & \cdots & f_n(x_{i_1}) \\ f_1(x_{i_2}) & f_2(x_{i_2}) & \cdots & f_n(x_{i_2}) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_{i_{n+1}}) & f_2(x_{i_{n+1}}) & \cdots & f_n(x_{i_{n+1}}) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} f_0(x_{i_1}) \\ f_0(x_{i_2}) \\ \vdots \\ f_0(x_{i_{n+1}}) \end{pmatrix}$$

is inconsistent for every choice of $x_{i_1}, \dots, x_{i_{n+1}}$ from x_1, \dots, x_N . If we define, for every $B = \langle i_1, \dots, i_{n+1} \rangle \in [N]^{n+1}$,

$$G_B(x_1, \dots, x_N) = \det \begin{pmatrix} f_1(x_{i_1}) & f_2(x_{i_1}) & \cdots & f_n(x_{i_1}) & f_0(x_{i_1}) \\ f_1(x_{i_2}) & f_2(x_{i_2}) & \cdots & f_n(x_{i_2}) & f_0(x_{i_2}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_1(x_{i_{n+1}}) & f_2(x_{i_{n+1}}) & \cdots & f_n(x_{i_{n+1}}) & f_0(x_{i_{n+1}}) \end{pmatrix}$$

then condition (2) is equivalent to the requirement that $G_B(x_1, \dots, x_N) \neq 0$ for all $B \in [N]^{n+1}$. Note that for each choice of $B \in [N]^{n+1}$, $G_B^{-1}(0)$ has empty interior as a subset of X^N , as a consequence of Lemma 3.1.

To complete the argument, we must show that

$$Q := X^N \setminus \left(\bigcup_{B \in [N]^n} F_B^{-1}(0) \cup \bigcup_{B \in [N]^{n+1}} G_B^{-1}(0) \right)$$

is nonempty. But since X^N is Baire, Q is actually open and dense. Taking $q \in Q \subseteq X^N$, we see that $X_0 := \text{range}(q)$ suffices, by Floyd's lemma. \square

Note that in the special case in which f_0, \dots, f_n are real analytic functions defined on \mathbb{R}^k , the set Q as in the proof of the theorem is not only dense and open, but co-null in the sense of Lebesgue measure by Corollary 3.2. This gives applications to probability distributions which have the same null sets as Lebesgue measure. Recall that a measure ν defined on the Borel sets is *absolutely continuous* with respect to λ if $\lambda(B) = 0$ always implies $\nu(B) = 0$. It is known that the Gaussian measures are absolutely continuous with respect to Lebesgue measure [11]. The same is true for the uniform probability measure defined on a box in Euclidean space, since this is just Lebesgue measure normalized to a bounded set.

Corollary 3.3. *Suppose f_0, f_1, \dots, f_n are linearly independent real analytic functions defined on*

$$R = \underbrace{[0, 1] \times [0, 1] \times \dots \times [0, 1]}_{k \text{ times}}.$$

Put $\mathcal{C} = \text{pos}(f_0 - \text{span}\langle f_1, \dots, f_n \rangle)$. For $N > n$, let ν be a probability measure on R^N which is absolutely continuous with respect to Lebesgue measure. Then if $q \in R^N$ is selected at random according to ν , then $\mathcal{C}|_{X_0}$ is maximum of VC dimension n with probability 1, where $X_0 = \text{range}(q)$.

Proof. Observe that $q \in Q$ almost surely, because $\nu(Q) = \lambda(Q) = 1$. \square

Note that for any set of real variables $V = \{v_1, \dots, v_m\}$, distinct monomials arising from V are linearly independent and analytic. Thus the above results apply, in particular, to polynomial functions and their sets of positivity.

This generalizes the Floyd/Dudley result which states that the set of open balls in a Euclidean space has the maximum property on points in general position.

It also applies to some functions which seem not to have been considered before, such as trigonometric polynomials. That is, functions of the form

$$t(x, y; a_0, a_1, \dots, a_N, b_1, \dots, b_N) = a_0 + \sum_{n=1}^N a_n \cos(nx) + \sum_{n=1}^N b_n \sin(nx) - y$$

where the a_i and b_i are viewed as parameters. The Wronskian criterion [18] for the linear independence of functions can be used to generate still more examples.

The notion of samples which are dense (in the product topology) with certain properties has been undertaken by Sontag in the context of neural networks [19].

References

- [1] S. Ben-David, and A. Litman, *Combinatorial Variability of Vapnik-Chervonenkis Classes*. Discrete Applied Mathematics, **86**, 1 (1998) pp 3-25.
- [2] T. Cover, *Geometrical and statistical properties of systems of linear inequalities with applications to pattern recognition*, IEEE Trans. Electron. Comput. EC-14 (1965) pp. 326-334.
- [3] R.M. Dudley, *Balls in \mathbb{R}^k do not cut all subsets of $k + 2$ points*, Adv. in Math. 31, 3 (1978) pp. 306-308.

- [4] R.M. Dudley, *Uniform central limit theorems*, Cambridge University Press, New York, 1999.
- [5] H. Federer, *Geometric Measure Theory*, Springer, Berlin, 1996.
- [6] S. Floyd, *Space-bounded learning and the Vapnik-Chervonenkis dimension*, Ph.D. thesis, U.C. Berkeley, 1989.
<http://www.icsi.berkeley.edu/pubs/techreports/tr-89-61.pdf>
- [7] S. Floyd, M.K. Warmuth, *Sample compression, learnability and the Vapnik-Chervonenkis dimension*, Machine learning, **21**,3,(1995),pp 269-304.
- [8] Vincent Guingona and Cameron Donnay Hill, *Local dp-rank and VC-density over indiscernible sequences*, preprint arXiv:1108.2554.
- [9] H. Johnson, *Dp-rank and forbidden configurations*, Notre Dame Journal of Formal Logic, **54**, 1 (2013), pp. 1–15.
- [10] S. Krantz, H. Parks, *A Primer of Real Analytic Functions*, 2nd Ed., Birkhäuser, New York, 2002.
- [11] Athreya, Krishna B.; Lahiri, Soumendra N. *Measure theory and probability theory*, Springer, New York, 2006.
- [12] D. Kuzmin, M. K. Warmuth, *Unlabeled Compression Schemes for Maximum Classes*, Journal of Machine Learning Research, **8** (2007), pp. 2047-2081.
- [13] Ulrike von Luxburg, Olivier Bousquet, and Bernard Schölkopf, *A compression approach to support vector model selection*, Journal of Machine Learning Research, 5(Apr), (2004), 293–323.
- [14] H. Royden, *Real Analysis*, 3rd Ed., Macmillan, New York, 1988.
- [15] B. I. P. Rubinstein, J. H. Rubinstein, *A Geometric Approach to Sample Compression*, Journal of Machine Learning Research, **13**, April (2012), pp. 1221-1261.
- [16] N. Sauer, *On the density of families of sets*, Journal of Combinatorial Theory, **13** (1972), pp 145–147.
- [17] S. Shelah, *A combinatorial problem: stability and order for models and theories in infinitary languages*, Pacific Journal of Mathematics **41**, 1 (1972), pp 247–261.
- [18] D. Somasundaram, *Ordinary Differential Equations: A First Course*, Alpha Science, 2001.
- [19] E. Sontag, *Shattering all sets of ‘k’ points in ‘general position’ requires (k-1)/2 parameters*, Neural Computation, 9, (1997), 337-348.
- [20] V. Vapnik and A. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory of Probability and its Applications **16** (1971), no. 2, 264–280.
- [21] E. Welzl, *Complete range spaces*, Unpublished notes, 1987.
- [22] R. S. Wenocur, R.M. Dudley, *Some special Vapnik-Chervonenkis classes*, Discrete Math., **33**, (1981),pp 313-318.