Machine Learning, 18, 131–148 (1995)

Bounding the Vapnik-Chervonenkis Dimension of Concept Classes Parameterized by Real Numbers

PAUL W. GOLDBERG pwgoldb@cs.sandia.gov

Department 1423, Sandia National Laboratories, MS 1110, P.O. Box 5800, Albuquerque, NM 87185-1110, U.S.A.

MARK R. JERRUM mrj@dcs.ed.ac.uk

Department of Computer Science, The University of Edinburgh, The King's Buildings, Mayfield Rd, Edinburgh EH9 3JZ, U.K.

Editor: Sally A. Goldman

Abstract. The Vapnik-Chervonenkis (V-C) dimension is an important combinatorial tool in the analysis of learning problems in the PAC framework. For polynomial learnability, we seek upper bounds on the V-C dimension that are polynomial in the syntactic complexity of concepts. Such upper bounds are automatic for discrete concept classes, but hitherto little has been known about what general conditions guarantee polynomial bounds on V-C dimension for classes in which concepts and examples are represented by tuples of real numbers. In this paper, we show that for two general kinds of concept class the V-C dimension is polynomially bounded in the number of real numbers used to define a problem instance. One is classes where the criterion for membership of an instance in a concept can be expressed as a formula (in the first-order theory of the reals) with fixed quantification depth and exponentially-bounded length, whose atomic predicates are polynomial inequalities of exponentially-bounded degree. The other is classes where containment of an instance in a concept is testable in polynomial time, assuming we may compute standard arithmetic operations on reals exactly in constant time.

Our results show that in the continuous case, as in the discrete, the real barrier to efficient learning in the Occam sense is complexity-theoretic and not information-theoretic. We present examples to show how these results apply to concept classes defined by geometrical figures and neural nets, and derive polynomial bounds on the V-C dimension for these classes.

Keywords: Concept learning, information theory, Vapnik-Chervonenkis dimension, Milnor's theorem

1. Introduction

This paper is concerned with inductive concept learning, and in particular, bounds on the sample size required to guarantee that a consistent hypothesis drawn from a given class will achieve reliable generalization. We work within the Probably Approximately Correct (PAC) framework introduced by Valiant (1984,1985).

Where concepts are discrete entities (representable using words over a finite alphabet) the results of Blumer et al. (1987) provide bounds in terms of the description length of the hypothesis. More generally, Blumer et al. (1989) show related bounds on the sample size required in terms of the *Vapnik-Chervonenkis dimension* of a concept class. Consequently it is of interest to establish bounds on the Vapnik-Chervonenkis dimension of non-discrete concept classes, which is what we are concerned with here.

This paper is organized as follows. In section 1.1 we define the learning model and consider the theoretical background of this work. In section 1.2 we discuss the significance of our results, their relationship to previous work, and their applicability. Section 2.1 gives the main theorems and proofs for upper bound results, and in section 2.2 we identify lower bounds by considering specific concept classes. In section 3 we consider two application areas, namely neural networks and geometrical learning problems.

1.1. Preliminary Definitions

In this paper, we consider the Vapnik-Chervonenkis dimension as it applies to the distribution-free PAC learning model. We give here the basic definitions involved; readers interested in the main results of the theory are referred to recent textbooks such as those by Natarajan (1991) and Anthony and Biggs (1992) which provide this background.

In PAC learning, the set of all objects that may be presented to the learner is called the *instance domain*, usually denoted X. Members of X (instances) are classified according to membership or non-membership of an unknown subset C of X, called the *target concept*, and the goal of the learner is to construct a hypothesis H that is a good approximation to C. The target concept C is restricted to be a member of a known collection $\mathcal{C}$ of subsets of X, called the *concept class*. Examples are assumed to be generated according to a fixed but unknown probability distribution P on X; we say that hypothesis H *ϵ-approximates* concept C, if the probability that H and C disagree on a random instance drawn according to P is at most ϵ. The criterion for successful learning is that the hypothesis should reliably classify further instances drawn according to P. This criterion is captured in the notion of "learning function." A *learning function* takes positive real parameters ϵ and δ (representing error and uncertainty, respectively) and a sequence of classified instances of the target concept drawn from the distribution P, and produces a hypothesis that ϵ-approximates the target concept with probability at least $1 - \delta$.

A learning algorithm — i.e., a computational procedure that implements a learning function — should ideally run in time polynomial in ϵ^{-1} and δ^{-1}, and the other parameters of the learning problem. In this paper however, we are not interested in complexity-theoretic considerations, but rather on the sample size required for a computationally unbounded learner to have enough information to be sure that a consistent hypothesis is reliable. Blumer et al. (1989) showed that the sample size required in this sense is directly proportional to the Vapnik-Chervonenkis dimension of the concept class under consideration. The precise statement of their main result is given in the next section.

The Vapnik-Chervonenkis dimension (which we will subsequently abbreviate to V-C dimension) was developed as a statistical tool by Vapnik and Chervonenkis (1971). It is defined as follows.

Definition. Let X be a set and let $\mathcal{C}$ be a collection of subsets of X (thus $\mathcal{C} \subseteq 2^X$). We say that a subset $S \subseteq X$ is *shattered* by $\mathcal{C}$ if for every partition of S into disjoint subsets S_1 and S_2, there exists $C \in \mathcal{C}$ such that $S_1 \subseteq C$ and $S_2 \cap C = \emptyset$. Then the *Vapnik-Chervonenkis dimension* of $\mathcal{C}$ is the maximum cardinality of any subset of X

shattered by C, or ∞ if arbitrarily large subsets may be shattered. The V-C dimension of concept class C will be denoted $\mathrm{VCdim}(C)$.

1.2. Background

The V-C dimension was studied by Blumer et al. (1989) as a means of analyzing non-discrete concept classes, in particular classes (typically geometrically motivated) where concepts and/or instances are naturally represented using real numbers. (The assumption is that one unit is charged for representing and operating on a real number, which is used in the neural networks literature, and noted by Valiant (1991) to be typically appropriate for geometrical domains.) In these situations, one cannot use counting arguments to show that a consistent hypothesis of some limited complexity achieves PAC learning. However, Blumer, Ehrenfeucht, Haussler and Warmuth (1989) showed that the V-C dimension of a concept class determines how many examples are necessary for a learner to form a PAC hypothesis. The following theorem of Blumer et al. shows in particular that uniform learnability is characterized by finiteness of the V-C dimension. Here and throughout the paper, log denotes logarithm to base 2, and ln denotes natural logarithm.

THEOREM 1.1 *(Blumer et al., 1989) Let C be a non-trivial, well-behaved* [1] *concept class.*

i. There exists a learning function (not necessarily polynomial-time computable) mapping samples to hypotheses in C if and only if $\mathrm{VCdim}(C)$ *is finite.*

ii. If the V-C dimension of C is d, where $d < \infty$ then

a. for $0 < \epsilon < 1$ and sample size at least

$$\max\left(\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{13}{\epsilon}\right)$$

any function mapping such samples to a consistent hypothesis in C is a learning function for C (which may not be evaluatable in polynomial time), and

b. for $0 < \epsilon < \frac{1}{2}$ and sample size less than

$$\max\left(\frac{1-\epsilon}{\epsilon}\ln\frac{1}{\delta}, d(1-2(\epsilon(1-\delta)+\delta))\right)$$

there is no learning function from such samples to any hypothesis class.

We may note that in establishing bounds on sample complexity in terms of the V-C dimension, Ehrenfeucht et al. (1989) have improved the above lower bound to $\Omega((1/\epsilon)\ln(1/\delta)+d/\epsilon)$.

Our interest here is in polynomial learnability, which is achieved when a learning algorithm runs in time polynomial in certain parameters of the learning problem. Besides

ϵ^{-1} and δ^{-1}, these are the domain dimension and the syntactic complexity of the target concept, which we will denote by n and k respectively. (In the familiar concept classes defined by Boolean formulas, n would typically correspond to the number of variables in a formula, and k to its length in some encoding.) In a class with infinite V-C dimension, we may still have non-uniform polynomial learnability with respect to one or both of these parameters, as we shall see momentarily.

This idea is treated in detail in Blumer et al. (1989) and Linial et al. (1991), but here we content ourselves with a simple observation based on theorem 1.1. Suppose that the instance domain X is partitioned or "stratified" according to domain dimension n, so that X is the disjoint union $\bigcup_{n\geq 1} X_n$. Suppose also that the concept class is further stratified according to concept complexity k, so that $\mathcal{C} = \bigcup_{k,n\geq 1} \mathcal{C}_{k,n}$, where $\mathcal{C}_{k,n} \subseteq 2^{X_n}$, for every k. Even though the V-C dimension of $\mathcal{C}$ may be unbounded, theorem 1.1 assures us that PAC learning of a concept in $\mathcal{C}_{k,n}$ is possible from a learning set of size polynomial in k and n, provided that the V-C dimension of the $\mathcal{C}_{k,n}$ is bounded by a polynomial in k and n, and vice versa[2].

In short, if the V-C dimension of $\mathcal{C}_{k,n}$ does not grow too rapidly with k and n then there is no information-theoretic barrier to efficient PAC learning, though computational intractability of the consistent hypothesis problem may well provide a complexity-theoretic barrier. For this reason it is important to find good general techniques for bounding the V-C dimension of stratified concept classes, and that is our goal in this paper. In particular, we seek upper bounds on the V-C dimension, as a function of k and n. Whenever this bound is polynomial in k and n, we are assured that a polynomial-sized sample contains enough information for a consistent hypothesis to achieve PAC-ness, following our earlier discussion.

There have been many previous papers presenting results giving the V-C dimension of classes of objects parameterized using real numbers. One result of Wenocur and Dudley (1981) is that the V-C dimension of halfspaces of $\mathbb{R}^n$ is $n+1$. Blumer et al. (1989) give other examples. Related to our work is the paper of Ben-David and Lindenbaum (1993), which also uses the theorem of Milnor to obtain quantitative bounds on the V-C dimension of continuously parameterized concept classes. They obtain results (motivated by computer vision) bounding the V-C dimension of geometrical concept classes defined by the set of transformed images of a shape, and also defined by a description of a general set of shapes. They show that the bounds they obtain are applicable to distribution-specific learning.

1.3. Summary of Results

We consider general concept classes whose concepts and instances are represented by tuples of real numbers. For such a concept class $\mathcal{C}$, let $\mathcal{C}_{k,n}$ be $\mathcal{C}$ restricted to concepts represented by k real values and instances represented by n real values. (So subclasses are parameterized by both concept complexity and instance size.)

Definition. The *membership test* of a concept class $\mathcal{C}$ over domain X takes as input a concept $C \in \mathcal{C}$ and instance $a \in X$, and returns the boolean value $a \in C$.

The membership test of a concept class can be thought of as a formula or algorithm taking as input representations of a concept and instance, and evaluating to the boolean value indicating membership. For a concept class with concepts or instances of varying representational complexity, it is convenient to express the membership test as a family of formulas/algorithms indexed by the complexity of the concepts and instances that they may take as input. The membership test for $\mathcal{C}_{k,n}$ as defined above, is assumed to be expressed as a formula $\Phi_{k,n}$ (in the first-order theory of the reals) with $k+n$ free variables representing a concept C and instance a; or as an algorithm $\mathcal{A}_{k,n}$, similarly taking $k+n$ real inputs, which uses exact real arithmetic and returns the truth value $a \in C$. In this situation we say that $\mathcal{C}_{k,n}$ is *defined by* $\Phi_{k,n}$ or $\mathcal{A}_{k,n}$.

By way of example, let m be a positive integer, $k = m(n+1)$, and define $\Phi_{k,n}$ by

$$\Phi_{k,n} = \bigvee_{i=1}^{m} \left[\sum_{j=1}^{n} (x_j - a_{ij})^2 \leq r_i^2 \right].$$

This formula defines the concept class whose elements are unions of m balls in n-dimensional Euclidean space: the variables a_{ij} parameterize the centers of the m balls, r_i their radii, and x_j the Cartesian coordinates of the instance.

We seek general conditions on $\Phi_{k,n}$ or $\mathcal{A}_{k,n}$ which guarantee that $\mathrm{VCdim}(\mathcal{C}_{k,n})$ be polynomial in k and n. This approach is in the same spirit as the work of Laskowski (1992), who exhibits a necessary and sufficient condition on a first-order formula Φ over some structure to define a class of finite V-C dimension. Note that a number of authors, e.g., Dudley (1978) and Stengle and Yukich (1989), have considered concept classes similar to the ones treated here, and shown that the V-C dimension is always finite. However, it appears that no attempt was previously made to *quantify* the V-C dimension in terms of natural parameters of the defining formulas.

For classes of the form defined above, we have the following results:

1. For a hierarchy of concept classes $\{\mathcal{C}_{k,n}\}_{k,n \in \mathbb{N}}$ defined by formulas $\Phi_{k,n}$ in the first-order theory of the real numbers, having fixed quantification depth, exponentially-bounded (in k, n) length and atomic predicates of exponentially-bounded degree, the V-C dimension of $\mathcal{C}_{k,n}$ is polynomial in k, n.

2. For a hierarchy of concept classes $\{\mathcal{C}_{k,n}\}_{k,n \in \mathbb{N}}$ defined by algorithms $\mathcal{A}_{k,n}$ which run in time polynomial in k, n, the V-C dimension of $\mathcal{C}_{k,n}$ is also polynomial in k, n. The algorithm $\mathcal{A}_{k,n}$ is allowed to perform conditional jumps (conditioned on equality and inequality of real values) and execute the standard arithmetic operations on real numbers exactly in constant time.

3. The above results can be extended so that k and n are limits on the complexity of concepts and examples, rather than the exact complexity.

Precise statements of these results are in the next section. We conclude this section by considering their importance to the area of learning theory.

The results appear to cover almost all continuously parameterized concept classes one might reasonably wish to consider in the context of PAC learning. The conditions seem to

rule out only the use of certain non-algebraic functions such as $\lfloor \cdot \rfloor$ or sine, or "unnatural" encodings of an object in terms of real values, such as compressing several real numbers into one by interleaving their decimal expansions.

The results are easy to apply: efficient PAC learning for continuously parameterized concept classes is reduced, virtually automatically, to the complexity-theoretic problem of efficiently finding a hypothesis consistent with the learning set.

The results are independent of the way a real number is represented computationally. Consider by contrast the following fact:

Observation: Assume that real numbers are represented using d-digit binary expansions. Then $\mathrm{VCdim}\, C_{k,n} \leq kd$.

This follows from the fact that there are at most 2^{kd} distinct concepts in $C_{k,n}$ (for any n), hence they cannot shatter more than kd instances. This observation is a representation-dependent result, and does not depend on any features of the concept class. Hence it is unenlightening from the point of view of learnability considerations.

Our results highlight the importance of the form taken by the membership test on the information content of a sample. The second result noted above is particularly significant, in providing a polynomial guarantee on the V-C dimension for concept classes where recognition of membership of an instance in a concept can be done in polynomial time (in the given model of computation). This is rarely a consideration for concept classes considered in the literature, but it has been noted to be necessary for a concept classes to be predictable.

The bounds are tight, modulo multiplication by a constant. (Explicit examples for lower bounds are constructed in section 2.2.) Hence any upper bounds for particular concept classes having a lower order of growth will have to rely on the internal structure of the concept class, rather that just the properties identified above.

2. Main Results

Throughout this section, $C_{k,n}$ will be as defined in section 1.2. A concept C and instance a will be represented by the sequences of reals $(y_1, ..., y_k)$, and $(x_1, ..., x_n)$ respectively. The symbol e always denotes the base of the natural logarithm.

2.1. Upper Bounds

A key theorem that we will use is an upper bound of Warren (1968) on the number of consistent sign assignments to a set of multivariate polynomials. A sign assignment to polynomial p is one of the (in)equalities $p > 0$ or $p = 0$ or $p < 0$; a sign assignment to a set of m polynomials is *consistent* if all m (in)equalities can be satisfied simultaneously by some assignment of real numbers to the variables. A *non-zero* sign assignment to p is one of the inequalities $p > 0$ or $p < 0$. Warren's upper bound is the following:

THEOREM 2.1 *(Warren, 1968) If $\{p_1, ..., p_m\}$ is a set of polynomials of degree at most $d \geq 1$ in n real variables with $m \geq n$, then the number of consistent non-zero sign assignments to the p_i is at most $(4edm/n)^n$.*

We have the following corollary:

COROLLARY 2.1 *If $\{p_1, ..., p_m\}$ is a set of polynomials of degree at most $d \geq 1$ in n real variables with $m \geq n$, then the number of consistent sign assignments to the p_i is at most $(8edm/n)^n$.*

Proof: Let $\mathcal{P} = \{p_1, ..., p_m\}$. Consider the set of polynomials

$$\mathcal{P}' = \{p_1 + \epsilon, p_1 - \epsilon, ..., p_m + \epsilon, p_m - \epsilon\}.$$

We claim that for sufficiently small $\epsilon > 0$, every sign assignment to $\mathcal{P}$ corresponds to a unique non-zero sign assignment to $\mathcal{P}'$.

For each consistent sign assignment to $\mathcal{P}$, choose n real assignments to their variables that satisfy it. Then the p_i evaluate to real numbers, and $\epsilon > 0$ is chosen to be strictly less than the absolute value of any of these which are non-zero, for all consistent sign assignments.

Now it may readily be seen that the same collection of sets of values for the n variables yield distinct non-zero sign assignments to $\mathcal{P}'$. Hence applying theorem 2.1 to $\mathcal{P}'$ gives the required limit on the number of sign assignments to $\mathcal{P}$. ■

We may note in passing that a similar but weaker result follows from a theorem of Milnor (1964) which has been used in other works in complexity theory to establish upper and lower bounds. Milnor's theorem gives an upper bound on the number of connected components of the subset of $\mathbb{R}^n$ corresponding to any sign assignment. Ben-David and Lindenbaum (1993) use the Milnor theorem in obtaining their upper bounds on the V-C dimension. We show next that the V-C dimension is at most logarithmic in the number of consistent sign assignments for the polynomials involved in the membership test of a concept class, which leads to the applicability of the Milnor/Warren bounds. However, the Warren bound is more directly applicable since it is really sign assignments that we are concerned with. (Each sign assignment corresponds to one or more connected components, which is why an upper bound on number of connected components gives an upper bound on the number of sign assignments.)

In what follows, the *size* of a formula refers to the number of *distinct* atomic predicates that it contains.

THEOREM 2.2 *Let $\{\mathcal{C}_{k,n} : k, n \in \mathbb{N}\}$ be a family of concept classes where concepts in $\mathcal{C}_{k,n}$ and instances are represented by k and n real values, respectively. Suppose that the membership test for any instance a in any concept C of $\mathcal{C}_{k,n}$ can be expressed as a boolean formula $\Phi_{k,n}$ containing $s = s(k,n)$ distinct atomic predicates, each predicate being a polynomial inequality or equality over $k + n$ variables (representing C and a) of degree at most $d = d(k,n)$. Then* $\mathrm{VCdim}(\mathcal{C}_{k,n}) \leq 2k \log(8eds)$.

COROLLARY 2.2 *Let $\mathcal{C}_{k,n}$ be as in theorem 2.2. If the size s and degree d are both at most exponential in k and n, then the V-C dimension* $\mathrm{VCdim}(\mathcal{C}_{k,n})$ *is polynomially bounded in k and n.*

Example: Consider the "union of balls" concept class from the previous section. There, the concept class $\mathcal{C}_{k,n}$ is defined by a formula with m atomic predicates of degree 2 involving $m(n+1)$ concept parameters. From theorem 2.2, the V-C dimension of $\mathcal{C}_{k,n}$ is bounded by $2m(n+1)\log 16em$. It is not known if this is tight, but it is certainly good — we may readily identify $m(n+1)$ as a lower bound. □

We next prove theorem 2.2.

Proof: Assume that $\Phi = \Phi_{k,n}$ is a formula having free variables $\{x_1, ..., x_n\}$ representing an instance, and $\{y_1, ..., y_k\}$ representing a concept. Let $s = s(k,n)$ be the size of Φ; hence Φ contains s distinct polynomials. Let $d = d(k,n)$ be an upper bound on their degree.

Let $v = v(k,n)$ be the V-C dimension of $\mathcal{C}_{k,n}$, and suppose $\{a_1, ..., a_v\}$ is a shattered set of instances. For each a_i let $\Phi(a_i)$ denote the formula in $\{y_1, ..., y_k\}$ obtained by substituting the values a_i has for $\{x_1, ..., x_n\}$ in Φ. Let S be the union over $i = 1, ..., v$ of all polynomials contained in $\Phi(a_i)$. So $|S| \leq vs$.

For $\{a_1, ..., a_v\}$ to be shattered, the set S of polynomials must be able to take 2^v different sign assignments, for various values of $y_1, ..., y_k$. From corollary 2.1, the number of sign assignments is bounded above by $(8edm/k)^k$ where $m = vs$ is the number of polynomials, d is their degree, and k is the number of variables. Hence we have the upper bound $(8edvs/k)^k$ on the number of sign assignments. Thus, $2^v \leq (8edvs/k)^k$; equivalently, taking logs to base 2 on each side,

$$v \leq k\log(8edsv/k) = k\log(8eds) + k\log(v/k).$$

Consider the following two cases:

(i) Suppose $8eds \geq v/k$; then $v \leq 2k\log(8eds)$.

(ii) Otherwise, $v \leq 2k\log(v/k)$, implying $v \leq 4k$.

In either case, $v \leq 2k\log(8eds)$. ■

We next show an interesting development of the foregoing results, by extending them to concept classes whose membership tests are programs described by bounded-depth algebraic decision trees. A decision tree is a computation tree each of whose leaves gives one value of a binary-valued output. The runtime of a program in the model we describe becomes the depth of the associated decision tree. For various computational problems, Steele and Yao (1982) and Ben-Or (1983) have used Milnor's bound to give lower bounds on the depth of any algebraic computation tree that solves them.

THEOREM 2.3 *Let $\{\mathcal{C}_{k,n} : k, n \in \mathbb{N}\}$ be a set of concept classes as before, for which the test for membership of an instance a in a concept C consists of an algorithm $\mathcal{A}_{k,n}$ taking*

$k+n$ real inputs representing C and a, whose runtime is $t = t(k,n)$, and which returns the truth value $a \in C$. The algorithm $\mathcal{A}_{k,n}$ is allowed to perform conditional jumps (conditioned on equality and inequality of real values) and execute the standard arithmetic operations on real numbers $(+,-,\times,/)$ in constant time. Then $\mathrm{VCdim}(\mathcal{C}_{k,n}) = O(kt)$.

COROLLARY 2.3 *Let $\mathcal{C}_{k,n}$ be as in theorem 2.3. If the runtime of algorithm $\mathcal{A}_{k,n}$ is polynomially bounded in k and n, then so is the V-C dimension of the concept class $\mathcal{C}_{k,n}$.*

We next prove theorem 2.3.

Proof: We transform $\mathcal{A}_{k,n}$ into a formula $\Phi_{k,n}$ of the form in theorem 2.2.

The algorithm $\mathcal{A}_{k,n}$ takes $k+n$ real inputs, $\{y_1,...,y_k\}$ representing concept C, and $\{x_1,...,x_n\}$ representing instance a. We assume that each line in $\mathcal{A}_{k,n}$ is of the one of the possible forms:

(i) $v_i := v_j \circ v_k$,

where v_j, v_k are either inputs or previously computed values, and $\circ$ is one of the standard arithmetic operators.

(ii) if v_i ($>$ or $=$ or $<$) 0 then goto line L,

where v_i is an input or previously computed value, and L is the label of some line in $\mathcal{A}_{k,n}$.

(iii) output "True" or "False".

Let $t(k,n)$ be the runtime of $\mathcal{A}_{k,n}$. Then we claim that any value computed during the execution of $\mathcal{A}_{k,n}$ is a rational function of $\{y_1,...,y_k,x_1,...,x_n\}$, with degree bounded above by 2^t. (The degree of a rational function p/q, p and q polynomials, is the sum of the degrees of p and q. The degree of a value v_i computed during the execution of $\mathcal{A}_{k,n}$ is to be interpreted as its degree when viewed as a (minimum degree) rational function of the input values.)

Only steps of type (i) may generate values of higher degree than ones computed previously, and the new value v_i cannot have degree higher than the sum of the degrees of v_j and v_k, whatever arithmetic operation is being used. Hence the degree of a value computed by $\mathcal{A}_{k,n}$, expressed as a rational function of the variables defining C and a, can at most double at each step of execution. The claim follows.

The algorithm $\mathcal{A}_{k,n}$ can be expressed as an algebraic decision tree with $\leq 2^t$ leaves, where each node with one child is an instruction of type (i), each node with two children is a test in an instruction of type (ii), and each leaf is labeled "True" or "False" corresponding to an instruction of type (iii). Each input causes the execution of $\mathcal{A}_{k,n}$ to take some path through this tree. In order to take a particular path, the inputs must satisfy at most t tests, consisting of (in)equalities of values computed by $\mathcal{A}_{k,n}$. So the condition for taking a particular path is a conjunction of such (in)equalities. (Note that an (in)equality involving two rational functions can be re-expressed as a polynomial (in)equality without increasing the degree.)

The predicate $a \in C$ can be expressed as a disjunction over the paths π ending in "True", of the conditions for taking π. Hence the predicate $a \in C$ may be expressed as a DNF Boolean formula, containing at most 2^t distinct atomic predicates, each of degree at most 2^t. Applying theorem 2.2 we see that the V-C dimension is bounded above by $2k\log(8e \times 2^t \times 2^t)$, or $2k(2t + \log(8e))$, which is $O(kt)$ as required. ■

Remarks:

1. Suppose that we want to consider k and n as limits on the complexity of instances and concepts, rather than their actual complexity. Then we can extend the above corollary to this case. We can construct an algorithm $\mathcal{A}'_{k,n}$ that tests membership of any example of complexity $\leq n$ in any concept of complexity $\leq k$. This algorithm takes $k+n$ real inputs representing the concept and example, and two further numbers which give the complexities of the concept and example to be tested. $\mathcal{A}'_{k,n}$ checks these and calls the appropriate $\mathcal{A}_{k,n}$.

2. As mentioned in the introduction, the theorem no longer holds if we may compute the $\lfloor \cdot \rfloor$ function in constant time. Neither does it hold if trigonometric functions such as sine are added to the computational model. Indeed the simple concept class

$$\{\{x : \sin xy \geq 0\} : y \in \mathbb{R}\}$$

already has unbounded V-C dimension. Sontag (1992) has used elaborations of this counterexample to refute some conjectures concerning the V-C dimension of neural nets.

3. Since the $O(kt)$ bound is tight (see section 2.2) a non-polynomial time algorithm $\mathcal{A}_{k,n}$ may define a concept class with more than polynomial V-C dimension. This can be viewed intuitively as saying that the amount of information that can be extracted from a real value is proportion to the time taken.

4. Suppose the concepts $\mathcal{C}_{k,n}$ are themselves programs of the same syntactic form as the $\mathcal{A}_{k,n}$ in the statement of the theorem, that take as inputs representations of instances, and test them for membership. Then the V-C dimension of $\mathcal{C}_{k,n}$ is polynomial in the syntactic complexity of $\mathcal{C}_{k,n}$, and its runtime. This follows from theorem 2.3 by regarding $\mathcal{A}_{k,n}$ as a "universal" program that takes as input $\mathcal{C}_{k,n}$ and instance a, and runs $\mathcal{C}_{k,n}$ on a.

Corollary 2.2 may be extended to an enriched class of membership tests, namely to predicates expressible in the first-order theory of the reals, of exponential size and degree as before, but with quantification allowed, though only to a uniformly bounded level of alternation of quantifiers and polynomially many quantified (i.e., not free) variables. In proving this result, the basic idea is that we may use a quantifier-elimination procedure of Renegar (1992) to give us a quantifier-free formula of the original form.

Using the notation of Renegar (1992), a formula in the first-order theory of the reals has the general form:

$$(Q_1 x^{[1]} \in \mathbb{R}^{n_1}) \cdots (Q_\omega x^{[\omega]} \in \mathbb{R}^{n_\omega}) P(y, x^{[1]}, ..., x^{[\omega]})$$

where the Q_i are quantifiers, $x^{[i]}$ is a vector of n_i real quantified variables, and $y = (y_1, ..., y_l)$ is a vector of real free variables. P consists of a boolean formula $\mathbb{P}$ having m atomic predicates consisting of polynomial equalities or inequalities of degree bounded by d (whose variables are in y or the $x^{[i]}$.)

THEOREM 2.4 *(Renegar, 1992) There is a quantifier-elimination procedure which requires only* $(md)^{2^{O(\omega)} l \Pi_k n_k}$ *operations and* $(md)^{O(l+\Sigma_k n_k)}$ *calls to* $\mathbb{P}$.

The algorithm constructs a quantifier-free formula of the form

$$\bigvee_{i=1}^{I} \bigwedge_{j=1}^{J_i} (h_{ij}(y) \circ_{ij} 0),$$

where

$$I \leq (md)^{2^{O(\omega)} l \Pi_r n_r},$$

$$J_i \leq (md)^{2^{O(\omega)} \Pi_r n_r},$$

the degree of h_{ij} *is at most* $(md)^{2^{O(\omega)} \Pi_k n_k}$, *and* $\circ_{ij}$ *represents one of the symbols* $\{<, \leq, =, \neq, \geq, >\}$.

COROLLARY 2.4 *Let* $\{\mathcal{C}_{k,n} : k, n \in \mathbb{N}\}$ *be a set of concept classes as before. Suppose that the membership test of a given instance* a *in a given concept* C *can be expressed as a formula* $\Psi_{k,n}$ *in the first-order theory of the real numbers with* $k + n$ *free variables representing* C *and* a*; suppose further that the number of bound variables in polynomial in* k *and* n*, that the depth of alternation of quantifiers is uniformly bounded, and that the atomic predicates are bounded in number and degree by an exponential function of* k *and* n*. Then the V-C dimension of* $\mathcal{C}_{k,n}$ *is polynomial in* k *and* n.

Proof: We use the quantifier elimination scheme of Renegar (1992) to reduce $\Psi_{k,n}$ to a formula $\Phi_{k,n}$ in the form of theorem 2.2. Note that it is the form of the quantifier-free formula rather than the time taken to construct it which is of importance.

The bounds on I and J_i in theorem 2.4 show that the number of polynomials acting as atomic predicates is doubly exponential in the depth of quantifier alternation (which we require to be constant for a concept class) and exponential in the number of quantified variables and the number of free variables.

Let $m(k, n)$ be the number of distinct atomic predicates in $\Psi_{k,n}$, $d(k, n)$ be the degree of the polynomials therein, $b(k, n)$ the number of quantified variables, and ω the (constant) number of quantifiers.

Applying theorem 2.4, we find that $\Phi_{k,n}$ is a quantifier-free formula whose size (number of distinct atomic predicates) is at most the product of the I and J_i in theorem 2.4, that

is:

$$(md)^{2^{O(\omega)}l\,\Pi_k n_k} \times (md)^{2^{O(\omega)}\Pi_k n_k}.$$

and since $\Pi_r n_r \leq b(k,n)^\omega$, this quantity is:

$$(md)^{(1+l)(2^{O(\omega)}b^\omega)}.$$

The degree of polynomials appearing in $\Phi_{k,n}$ is $(md)^{2^{O(\omega)}b^\omega}$. Hence by theorem 2.2, the V-C dimension of the concept class thus defined is $O(k \log sd)$, where s and d are the number and the maximum degree of polynomials in $\Phi_{k,n}$, which is:

$$O\big(k(2+l)2^{O(\omega)}b^\omega \log md\big)$$

Recall that l is the number of free variables, i.e., $l = k + n$. The result follows. ■

2.2. *Lower Bounds*

The "union of balls" example raises the question of how close theorem 2.2 is to being best possible. Here, we show how to construct concept classes, parameterized by k, s, and d, whose V-C dimension is within a constant factor of the upper bound of theorem 2.2, provided only that $d \geq 2$. We work towards the general construction via some special cases.

First consider the situation $k = d = 1$, i.e., the concept class is parameterized by a single variable, and all the atomic formulas are linear. Let $v \in \mathbb{N}^+$, $s = 2^v - 1$, and $\alpha = a_0 a_1 \ldots a_{s+v-1} \in \{0,1\}^{s+v}$ be a (non-cyclic) de Bruijn sequence, that is, a binary word α of length $s+v$ in which each binary word of length v occurs precisely once as a subword; such words α exist for all choices of v (Hall, 1967, p. 92). Define

$$\phi_s(x,y) = \bigvee_{i:a_i=1} (x = y + i),$$

where x and y are, respectively, an instance- and a concept-variable; note that $\phi_s(x,y)$ contains at most s atomic formulas, since at least v of the symbols a_i must be 0. Consider the concept class $\mathcal{C}_s = \big\{\{x : \phi_s(x,y)\} : y \in \mathbb{R}\big\}$ defined by ϕ_s. Concepts in $\mathcal{C}_s$ are "projections" of the de Bruijn sequence α onto the real line, which may be translated to any desired position by setting the parameter y. (It will be sufficient here to allow y to range over the natural numbers only.) It is clear from the definition of de Bruijn sequence that the point set $\{1,2,\ldots,v\}$ is shattered by $\mathcal{C}_s$; thus the V-C dimension of $\mathcal{C}_s$ is at least $v = \log(s+1)$, where s is (an upper bound on) the number of atomic formulas in ϕ_s. This example demonstrates that bound given in theorem 2.2 is tight in the special case $k = d = 1$.

Next set $d = 2$, and allow $k \geq 2$ to be arbitrary. Letting s be as before, define

$$\Phi_{k,s}(\mathbf{x},\mathbf{y}) = \Phi_{ks}(x_1,\ldots,x_k,y_1,\ldots,y_k)$$

$$= \bigvee_{i:a_i=1} \Big[\sum_{j=1}^{k} x_j(x_j - y_j - i) = 0 \Big].$$

The concept class $\mathcal{C}_{k,s} = \{\{\mathbf{x} : \Phi_{k,s}(\mathbf{x}, \mathbf{y})\} : \mathbf{y} \in \mathbb{R}^k\}$ shatters the set of kv points (instances) of the form $(0, \ldots, 0, u, 0, \ldots, 0) \in \mathbb{R}^k$, where $u \in \{1, 2, \ldots, v\}$. The mechanism through which this is achieved is similar to that described in the case $k = 1$; here, the concept-variables y_j act independently to select an arbitrary subset of the v instances that have non-zero jth coordinate. Hence the V-C dimension of $\mathcal{C}_{ks}$ is at least $kv = k \log(s + 1)$. Thus theorem 2.2 is also tight in the special case $d = 2$.

Analogously, when $s = 1$ but $k > 1$, we may set $d = 2(2^v - 1)$ and define

$$\Phi'_{k,d}(\mathbf{x}, \mathbf{y}) = \Phi'_{k,d}(x_1, \ldots, x_k, y_1, \ldots, y_k)$$

$$= \prod_{i:a_i=1} \Big[\sum_{j=1}^{k} x_j(x_j - y_j - 1) \Big] = 0,$$

which formula is equivalent to $\Phi_{k,s}$ and hence defines a concept class $\mathcal{C}'_{k,d}$ of V-C dimension at least $kv = \Theta(k \log d)$. Finally note that the constructions used for $\Phi_{k,s}$ and $\Phi'_{k,d}$ may be combined to provide a tradeoff between s and d. Let $2(2^v - 1) \leq sd < 2(2^{v+1} - 1)$, with d even. Factors with the general form $x_j(x_j - y_j - i)$ — where $1 \leq j \leq k$, and i satisfies $a_i = 1$ — may be multiplied together in groups of $d/2$ to yield (at most) s atomic formulas of degree d; the disjunction of these atomic formulas then yields a formula $\Phi_{k,s,d}$ defining a concept class $\mathcal{C}_{k,s,d}$ with V-C dimension at least $kv = \Theta(k \log sd)$. We thus have:

THEOREM 2.5 *For $k, s, d \in \mathbb{N}^+$ with $d \geq 2$, there exists a formula $\Phi_{k,s,d}$, over instance-variables $x_1, \ldots, x_k$ and concept-variables $y_1, \ldots, y_k$, and containing s atomic formulas of degree d, such that the concept class $\mathcal{C}_{ksd}$ defined by $\Phi_{k,s,d}$ has V-C dimension $\Omega(k \log sd)$.*

Note that that the above theorem demonstrates that theorem 2.2 is uniformly best possible, except perhaps when $d = 1$. The parameter n which denotes the number of instance-variables has been conspicuously absent from our deliberations, and it is possible that more refined bounds involving that parameter could be derived.

The following concept class shows that the upper bound of theorem 2.3 is also tight, modulo big-O notation.

We define a parameterized concept class $\mathcal{C}_{k,t}$ consisting of concepts defined by a sequence of k real values, with instances defined by one real value, and a membership testing algorithm with runtime t. Let a concept $C \in \mathcal{C}_{k,t}$ be represented by real numbers $y_1, \ldots, y_k$, and an instance by a real number x. The following algorithm shatters instances represented by numbers of the form 2^{-i}, for $i = 1, \ldots, kt$. The shattering set is the set of all concepts represented by numbers between 0 and 1 with binary expansions of length t, i.e., numbers of the form $j2^{-t}$, where j is an integer between 0 and 2^t.

For a concept C represented by $y_1, ..., y_k$, let

$$S_C = \sum_{i=1}^{k} \frac{y_i}{2^{t(i-1)}}.$$

Hence S_C can be any binary expansion of length kt, and consists of the concatenation of the y_i's.

Consider x and S_C as binary expansions. Let C contain an instance represented by x if and only if S_C has a 1 in the same position as the most significant 1 in x. (For elements of our shattered set, that is the only 1 in the expansion.) We now have to show that this test can be done using an $O(t)$ algorithm. The following search procedure works by first finding the block where the most significant bit of x is, and then we search in this block (i.e. y_i).

```
1.      Compute 2^t by repeated squaring.                    /*time O(log t)*/
2.      i := 1
3.      while (x < 2^{-t} and i < k) do
4.          x := 2^t · x;   i := i + 1;   od;                /*time O(k)*/
5.  /* Now we see if x and y_i have a common 1 within the first
6.  t positions:*/
7.      for j := 1 to t
8.          x := 2x;   y_i := 2y_i;
9.          If x ≥ 1 and y_i ≥ 1 then output "yes" and halt;
10.         If x ≥ 1 then x := x − 1;
11.         If y_i ≥ 1 then y_i := y_i − 1;
12.     next;                                                /* time O(t) */
13.     Output "no".
```

Note that t is not completely independent of k and n: we assume that $t \geq k + n$, which allows the membership testing algorithm time to read in the values representing the concept and instance. Observe that if we represent a concept by the single number S_C then (from the theorem) we cannot shatter the instances so quickly: the process of extracting the relevant bit in S_C apparently involves too many subtractions in order to extract (say) y_i from S_C.

To summarize, we have

THEOREM 2.6 *For $k, n, t \in \mathbb{N}^+$ with $t > k$ there exists an algorithm $\mathcal{A}_{k,n}$ with runtime t, which defines a concept class of V-C dimension $\Omega(kt)$.*

3. Applications

We consider briefly two kinds of concept class which can be seen to have polynomial V-C dimension as a result of the theorems in section 2.

3.1. *Neural Nets*

Let $\mathcal{C} = \bigcup_{k,n\in\mathbf{N}} \mathcal{C}_{k,n}$ be a concept class defined by a family of linear threshold, feed-forward neural net architectures with n inputs and k real weights. Baum and Haussler (1989) show that $\mathrm{VCdim}(\mathcal{C}_{k,n}) = O(k \log k)$. A fairly direct application of theorem 2.3 gives $\mathrm{VCdim}(\mathcal{C}_{k,n}) = O(k^2)$, an admittedly weaker result, but one obtained as a simple corollary to the general bound.

However, theorem 2.3 can be used for more than merely re-deriving old results. For example, the above result for linear thresholds can be generalized to many classes of networks with non-linear activation functions. One class of interest is networks which compute polynomial or piecewise polynomial functions of the inputs, instead of linear combinations. Provided these functions can be computed in time polynomial in the size of the network (using the usual arithmetic operations), then the output of the network can also be computed in polynomial time, and consequently, by corollary 2.2, such networks have V-C dimension polynomial in k and n. Polynomial upper bounds have recently been shown for bounded depth networks of this kind by Maass (1992); now we see that a similar result holds without the restriction to bounded depth. It appears likely that our results are applicable to other related classes of networks.

Note that our results *cannot* be applied to one of the standard activation functions, namely the "sigmoid" $1/(1 + \mathrm{e}^{-x})$. Using deep results from logic, Macintyre and Sontag (1992) have shown that the V-C dimension of such nets is finite, but no explicit bounds have been computed. However, theorem 2.3 does give a polynomial guarantee for sigmoid functions such as

$$\sigma(x) = \begin{cases} 1 - 1/(2x+2) \text{ if } x \geq 0 \\ 1/(2-2x) \text{ if } x \leq 0 \end{cases}$$

This function is differentiable and has the appropriate shape, and can be computed in the appropriate way required by theorem 2.3.

3.2. *Geometrical Classes*

We have seen that the V-C dimension is already known to be polynomial in concept and instance complexity for certain specific concept classes, such as halfspaces or balls in Euclidean space. There are however natural concept classes whose concepts are apparently more complex, and for which it may not be feasible to evaluate the V-C dimension exactly. In this section we look briefly at an example for which the theorems of section 2 provide a useful polynomial limit on the growth of the V-C dimension.

In Goldberg (1992) concept classes are considered whose concepts are sets of geometrical figures which are in some sense mutually resemblant. The associated learning problems are motivated by the problem of learning to recognize an object from a visual image of it. The learning examples are visual images, and a positive example is assumed to be "close" to some "ideal" image representing the object to be recognized. This is

made formal by defining some metric on geometrical figures with the property that two figures are close together under the metric if they are in fact similar in some sense. Then a concept can be defined as a sphere in this metric space.

For example, the Hausdorff metric is defined as follows. Let P_1, P_2 be two sets of points in a metric space (S, d), where d is a metric on elements of S. Then the Hausdorff distance between P_1 and P_2 is

$$H(P_1, P_2) = \max\left\{ \sup_{p_1 \in P_1} \left\{ \inf_{p_2 \in P_2} \{d(p_1, p_2)\} \right\}, \sup_{p_2 \in P_2} \left\{ \inf_{p_1 \in P_1} \{d(p_1, p_2)\} \right\} \right\}$$

So we would take $S = \mathbb{R}^2$, assuming that we are considering planar geometrical figures, with d the Euclidean distance between two points. The reason why the Hausdorff metric reflects geometrical resemblance is that for the distance between two sets of points in the plane (such as polygons) to be $\leq r$ we need every point on each set to be within r of some point on the other set.

If the geometrical figures we are considering are planar polygons, then an instance consisting of an n-gon is naturally represented by the $2n$ coordinates of its vertices. The concept of all polygons within distance r of some polygon P may be represented by the coordinates of the vertices of P and the value of r. This class should not be confused with ones where concepts are polygons (or polyhedra) and instances are points contained in them. An instance here is itself a polygon, one that is close under the Hausdorff metric to the concept polygon. Note that instance complexity is independent of concept complexity, since one can test proximity of polygons with different numbers of sides. The Hausdorff metric can be computed on polygons in polynomial time, using the algorithm of Alt, Behrends and Blömer (1991). In fact it can be computed in linear time for convex polygons (Atallah, 1983). This implies that the membership test for the concept class under consideration can be done in polynomial time, allowing us to apply corollary 2.3.

Alt et al. (1991) show that the Hausdorff metric remains polynomial-time computable when it is minimized over classes of isometries in the plane. This would correspond to a learning problem where the aim is to learn "shape" rather than "shape + position". Corollary 2.3 can then be applied in this case. Alternatively, the membership test for the concept class could be naturally expressed as a formula of the form indicated in corollary 2.4, with quantification over variables denoting position (which was the argument used in Goldberg (1992) to show that the V-C dimension is polynomial in this case).

4. Open Problems

A question of theoretical interest arising from corollary 2.4 is whether in fact the V-C dimension would still be polynomial if the number of quantifier alternations may be polynomial in n and k. It would be interesting to know if that answer is affected if $\Psi_{k,n}$ is required to have polynomial length. An optimal quantifier elimination scheme is not known, and indeed quantifier elimination elimination in conjunction with the upper bound of theorem 2.2 may not yield the best bound on the growth of the V-C dimension.

It would be interesting to obtain a generalization of theorem 2.3 that allowed a membership testing algorithm to evaluate e raised to the power of a computed value, in unit time. This would provide an upper bound for the V-C dimension for neural nets that compute the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$.

Acknowledgments

The second author is grateful to Eduardo Sontag for useful and encouraging discussions.

This work was performed by the first author at Sandia National Laboratories and was supported by the U.S. Department of Energy under contract DE-AC04-76DP00789. It was begun while he was supported by a graduate studentship of the UK Science and Engineering Research Council.

The second author was supported by grant GR/F 90363 of the UK Science and Engineering Research Council and by Esprit Working Group "RAND," and was partly done while he was visiting the NEC Research Institute at Princeton NJ, USA.

Notes

1. This is a relatively benign measure-theoretic assumption discussed in an appendix in Blumer et al. (1989).
2. It is not in fact necessary to restrict the search for a hypothesis to the same "stratum" as the target concept; it may be computationally advantageous to use as hypothesis class a higher stratum k, provided k is not too large. Blumer et al. (1987) formalize this notion in their definition of "Occam algorithm".

References

Alt, H., Behrends, B., & Blömer, J. (1991). Approximate Matching of Polygonal Shapes. *Procs. of the 1991 ACM Symposium on Computational Geometry*, pp. 186-193.

Anthony, M., & Biggs, N. (1992). *Computational Learning Theory: an Introduction*, Cambridge University Press, 1992.

Attalah, M.J. (1983). A Linear Time Algorithm for the Hausdorff-distance between Convex Polygons. *Information Processing Letters* **17** pp. 207-209.

Baum, E.B., & Haussler, D. (1989). What Size Net Gives Valid Generalization? *Neural Computation* **1**, pp. 151-160.

Ben-David, S., & Lindenbaum, M. (1993). Localization vs. Identification of Semi-Algebraic Sets. *Proceedings of the 6th Annual ACM Conference on Computational Learning Theory*, pp. 327-336.

Ben-Or, M. (1983). Lower Bounds for Algebraic Computation Trees. *Proceedings of the 15th Annual ACM Symposium on the Theory of Computing*, pp. 80-86.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M.K. (1987). Occam's Razor. *Information Processing Letters* **24** pp. 377-380.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M.K. (1989). Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the Association for Computing Machinery*, **36** No. 4, pp. 929-965.

Dudley, R.M. (1978). Central Limit Theorems for Empirical Measures, *Annals of Probability* **6**, pp. 899–929.

Ehrenfeucht, A., Haussler, D., Kearns, M., & Valiant, L.G. (1989). A General Lower Bound on the Number of Examples Needed for Learning. *Information and Computation* **82**, pp. 247-261.

Goldberg, P. (1992). PAC-Learning Geometrical Figures. *PhD thesis, Department of Computer Science, University of Edinburgh* (1992).

Hall, M. (1967). *Combinatorial Theory*, Blaisdell, Waltham MA (1967).

Haussler, D., Littlestone, N., & Warmuth, M.K. (1988). Predicting $\{0,1\}$ functions on randomly drawn points. *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science,* pp. 100-109.

Laskowski, M.C. (1992). Vapnik-Chervonenkis Classes of Definable Sets. *J. London Math. Society,* (2) 45, pp. 377-384.

Linial, N., Mansour, Y., & Rivest, R. (1991). Results on Learnability and the Vapnik-Chervonenkis Dimension. *Information and Computation* **90**, pp. 33-49.

Maass, W. (1992). Bounds for the Computational Power and Learning Complexity of Analog Neural Nets. *Insts. for Information Processing Graz, report 349; Oct. 1992. Proceedings of the 25th Annual ACM Symposium on the Theory of Computing (1993),* pp. 335-344.

Macintyre, A. & Sontag, E.D. (1993). Finiteness Results for Sigmoidal "Neural" Networks, *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing*, pp. 325-334.

Milnor, J. (1964). On the Betti Numbers of Real Varieties. *Procs. of the American Mathematical Society,* **15**, pp. 275-280.

Natarajan, B.K. (1991) *Machine Learning: A Theoretical Approach.* Morgan Kaufman Publishers, Inc., ISBN 1-55860-148-1

Renegar, J. (1992). On the Computational Complexity and Geometry of the First-Order Theory of the Reals. Part 1 (of 3). *Journal of Symbolic Computation* **13**, pp. 255-299.

Sontag, E.D. (1992). Feedforward Nets for Interpolation and Classification, *Journal of Computer and System Sciences* **45**, pp. 20–48.

Steele, J.M. & Yao, A.C. (1982). Lower Bounds for Algebraic Decision Trees. *Journal of Algorithms* **3**, pp. 1-8.

Stengle, G., & Yukich, J.E. (1989). Some New Vapnik-Chervonenkis Classes, Annals of Statistics **17**, pp. 1441–1446.

Valiant, L.G. (1984). A Theory of the Learnable. *Communications of the ACM,* **27** No. 11, pp. 1134-1142.

Valiant, L.G. (1985). Learning Disjunctions of Conjunctions. *Procs of the 9th International Joint Conference on AI,* pp. 560-566.

Valiant, L.G. (1991). A View of Computational Learning Theory. *NEC Research Symposium: Computation and Cognition (ed. C.W. Gear), SIAM, Philadelphia, 1991.*

Vapnik, V.N., & Chervonenkis, A.Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications* **16**, No. 2 pp. 264-280.

Warren, H.E. (1968). Lower Bounds for Approximation by Non-linear Manifolds. *Trans. of the AMS* **133**, pp. 167-178.

Wenocur, R.S., & Dudley, R.M. (1981). Some special Vapnik-Chervonenkis classes. *Discrete Mathematics* **33**, pp. 313-318.

Received October 3, 1993
Accepted May 5, 1994
Final Manuscript June 17, 1994