VC Dimension and Learnability of Sparse Polynomials and Rational Functions

Marek Karpinski * Thorsten Werther †

Abstract

We prove upper and lower bounds on the VC dimension of sparse univariate polynomials over reals, and apply these results to prove uniform learnability of sparse polynomials and rational functions. As another application we solve an open problem of Vapnik ([Vapnik 82]) on uniform approximation of the general regression functions, a central problem of computational statistics (cf. [Vapnik 82]), p. 256).

*Department of Computer Science, University of Bonn, and International Computer Science Institute, Berkeley, California. Supported in part by Leibniz Center for Research in Computer Science, by the DFG Grant KA 673/2-1, and by the SERC Grant GR-E 68297.

†Department of Computer Science, University of Bonn, and International Computer Science Institute, Berkeley, California.

1 Introduction

The paper studies the problem of computational identification (learnability) of sparse real polynomials and rational functions.

In [Val 84], Valiant introduced a model of learning concepts from examples taken from an unknown distribution. In this model, a concept c from a class C is a subset of an instance space X. Let C be a class of concepts from X. A labeled example $< x, + >$ or $< x, - >$ of a target concept c is an element of X, along with a label '+' or '$-$', indicating whether x is a member of the set c. In the Valiant model of learning, each example is drawn independently from a fixed but unknown distribution P on X. Each example is labeled either as a positive or as a negative example, consistently with the unknown target concept c.

The goal of the learning algorithm is to compute a good uniform approximation of the target concept, with high probability. Upper and lower bounds on the sample complexity for learning various concept classes have been given in [VC 71], [BEHW 87], [Fl 89]. These bounds are based on the Vapnik-Chervonenkis (VC) dimension of a class C.

For the corresponding problem of interpolation of polynomials over fields of characteristic zero cf. [GK 87], [BT 88] and over finite fileds cf. [GKS 88].

Definition 1. For a concept class C on X and for $S \subset X$, let $\Pi_C(S)$ be the set of subsets T of S such that $T = S \cap c$ for some concept c in C. Thus $\Pi_C(S)$ is the restriction of concept class C to the set S. If $\Pi_C(S) = 2^S$, then the set S is *shattered* by C. The *Vapnik-Chervonenkis dimension* (VC dimension) of the class C is the largest integer d such that some set $S \subset X$ of size d is shattered by C.

This paper explores the VC dimension of the concept class $\mathcal{P}_t$ consisting of the t-sparse polynomials over the real numbers, i.e

$$\mathcal{P}_t = \{\ f \mid f \in \mathbb{R}[x],\ f \text{ is } t\text{-sparse}\ \}.$$

Valiant's model of learning can be thought of as learning the border between positive and negative examples. In this situation we consider $\{0,1\}$-valued indicator functions ([Vap 82]). Hence, in the context of the Problem of Pattern Recognition [Vap 82], we define examples (x, y) from the instance space $X = (\mathbb{R}, \mathbb{R})$ to be labeled positive if the point (x, y) lies 'above' the t-sparse polynomial f ($f \in \mathcal{P}_t$ the unknown target concept), and vice versa, i.e.

$$< (x,y), + > \iff y \geq f(x) \qquad \text{and} \qquad < (x,y), - > \iff y < f(x).$$

Let $S \subset (\mathbb{R}, \mathbb{R})$ be the set of points $\{(x_i, y_i)\}_{i=1,\ldots,d}$ of size d for $x_1 < x_2 < \ldots < x_d$. A t-sparse polynomial f is said to satisfy a labeling $\sigma \in \{+,-\}^d$ on S if the points (x_i, y_i) are positive examples for f if $\sigma(i) = +$, and negative examples for f if $\sigma(i) = -$. The set S is shattered by the class of t-sparse polynomials $\mathcal{P}_t$, iff for each labeling $\sigma \in \{+,-\}^d$ there exists a t-sparse polynomial f_σ satisfying σ on S. We denote the VC dimension of the set $\mathcal{P}_t$ in this context by $\mathrm{VC}_{\geq}(\mathcal{P}_t)$.

Several generalizations of the standard PAC-model have been considered (cf. [Ha 89], [Vap 89]) in order to deal with real-valued functions instead with the indicator functions implied by the class $\mathcal{P}_t$ as above. We will use the notion proposed in [Ha 89].

Here for each $f \in \mathcal{P}_t$, an indicator function $I(f)$ is defined by

$$I(f)(x,y,\epsilon) = \begin{cases} 1 & \text{if } |f(x)-y| \leq \epsilon \\ 0 & \textit{otherwise} \end{cases},$$

where ϵ is any positive real number.

We define $\epsilon\text{-VC}(\mathcal{P}_t) = \mathrm{VC}(I(\mathcal{P}_t))$. Examples (x,y) from the instance space $X = (I\!R, I\!R)$ are labeled positive if the point (x,y) lies within ϵ-distance from the unknown target function, and vice versa.

This paper is organized as follows.

Section 2 gives lower and upper bounds on the $\mathrm{VC}_{\geq}$ dimension of sparse polynomials, proving that the class of sparse polynomials is uniformly learnable (cf. [BEHW 87]). Furthermore, upper and lower bounds on the sample complexity for learning sparse polynomials can be derived ([Fl 89], [Ha 89]). The more complete discussion of underlying algorithms and their analysis will be given in a full version of this paper.

Section 3 generalizes these bounds for the ϵ-VC dimension of t-sparse polynomials. Applying results of Vapnik ([Vap 82]), we prove the existence of algorithms that approximate the regression function uniformly by sparse polynomials.

2 Lower and Upper Bounds on $\mathrm{VC}_{\geq}(\mathcal{P}_t)$

2.1 Lower bounds on $\mathrm{VC}_{\geq}(\mathcal{P}_t)$

We start with a lower bound on the $\mathrm{VC}_{\geq}$ dimension of 1-sparse polynomials.

Lemma 2. *The $\mathrm{VC}_{\geq}$ dimension of 1-sparse polynomials is at least 3.*

Proof. We show that for each labeling $\sigma \in \{+,-\}^3$ there is a 1-sparse polynomial f_σ satisfying σ on the set $S = \{(-1,0),(1,2),(4,14)\}$ of size 3. Choose for example

σ	f_σ	σ	f_σ
$- - -$	15	$+ - -$	$5x$
$- - +$	3	$+ - +$	$3x$
$- + -$	x^2	$+ + -$	x^3
$- + +$	1	$+ + +$	-1

□

Remark 3. Let a set S of size d be shattered by the class of t-sparse polynomials. Then there is a set $Z = \{(x_i, y_i)\}_{i=1,\ldots,d}$ and constants $\epsilon_i > 0$, $i = 1, \ldots, d$, such that every set $S' = \{ \ (\bar{x}_i, \bar{y}_i) \mid |(\bar{x}_i, \bar{y}_i) - (x_i, y_i)| \leq \epsilon_i \ \}_{i=1,\ldots,d}$ is shattered by t-sparse polynomials.

PROOF. For each $\sigma \in \{+, -\}^d$ there is a t-sparse polynomial f_σ satisfying σ on S. For $i = 1, \ldots, d$ we define the regions

$$M_i = \{ \ (x, y) \mid \forall \sigma \in \{+, -\}^d : \left\{ \begin{array}{ll} y \geq f_\sigma & \text{if } \sigma(i) = + \\ y < f_\sigma & \text{if } \sigma(i) = - \end{array} \right\} \}.$$

Since S is shattered by $\{f_\sigma\}_{\sigma \in \{+,-\}^d}$ there exists a point (x_i, y_i) and a constant $\epsilon_i > 0$ such that the ball

$$B_{\epsilon_i}(x_i, y_i) = \{ \ (x, y) \mid |(x, y) - (x_i, y_i)| \leq \epsilon_i \ \}$$

is a prober subset of M_i, hence every set S' defined as above is shattered by the t-sparse polynomials $\{f_\sigma\}_{\sigma \in \{+,-\}^d}$. □

Given a set shattered by t_1-sparse polynomials and a set shattered by t_2-sparse polynomials we construct a set shattered by $(t_1 + t_2)$-sparse polynomials.

Lemma 4. $\mathrm{VC}_{\geq}(\mathcal{P}_{t_1+t_2}) \geq \mathrm{VC}_{\geq}(\mathcal{P}_{t_1}) + \mathrm{VC}_{\geq}(\mathcal{P}_{t_2})$.

PROOF. Let $d_1 = \mathrm{VC}_{\geq}(\mathcal{P}_{t_1})$ and $d_2 = \mathrm{VC}_{\geq}(\mathcal{P}_{t_2})$. Then there are sets of points S_1 and S_2 of size d_1, d_2 that are shattered by t_1-sparse polynomials and t_2-sparse polynomials, resp..

Let $S_1 = \{(x_i^{(1)}, y_i^{(1)})\}_{i=1,\ldots,d_1}$ and $S_2 = \{(x_j^{(2)}, y_j^{(2)})\}_{j=1,\ldots,d_2}$. For a labeling $\sigma^{(1)} \in \{+, -\}^{d_1}$ let $f_{\sigma^{(1)}}$ satisfy $\sigma^{(1)}$ on S_1, and for a labeling $\sigma^{(2)} \in \{+, -\}^{d_2}$ let $g_{\sigma^{(2)}}$ satisfy $\sigma^{(2)}$ on S_2.

In order to show that $\mathrm{VC}_{\geq}(\mathcal{P}_{t_1+t_2}) \geq d_1 + d_2$, we modify the sets S_1 and S_2 (and the corresponding polynomials shattering S_1 and S_2), such that the union of these modified sets is shattered by polynomials derived by adding some of the modified polynomials.

First, we pull the sets S_1, S_2 apart, such that the points in S_1 have an absolute x-coordinate of at most $1/2$ and the points in S_2 have an absolute x-coordinate of at least 2.

So let

$$c_1 > 2 \cdot \max_{(x_i,y_i)\in S_1}\{|x_i|\} \qquad \text{and} \qquad c_2 < \frac{1}{2} \cdot \min_{(x_j,y_j)\in S_2}\{|x_j|\}.$$

By Remark 3, we can assume that $c_2 > 0$.

Then the set

$$\bar{S}_1 = \{(\bar{x}_i, \bar{y}_i)\}_{i=1,\ldots,d_1} \quad \text{with } (\bar{x}_i, \bar{y}_i) = (\frac{1}{c_1}x_i, y_i),\ (x_i, y_i) \in S_1$$

is of size d_1 and is shattered by the set of t_1-sparse polynomials $\{\bar{f}_{\sigma^{(1)}}\}_{\sigma^{(1)}\in\{+,-\}^{d_1}}$, where $\bar{f}_{\sigma^{(1)}}(x) = f_{\sigma^{(1)}}(c_1 x)$.

Similarly the set

$$\bar{S}_2 = \{(\bar{x}_j, \bar{y}_j)\}_{j=1,\ldots,d_2} \quad \text{with } (\bar{x}_j, \bar{y}_j) = (c_2 x_j, y_j),\ (x_j, y_j) \in S_2$$

is of size d_2 and is shattered by the set of t_2-sparse polynomials $\{\bar{g}_{\sigma^{(2)}}\}_{\sigma^{(2)}\in\{+,-\}^{d_2}}$, where $\bar{g}_{\sigma^{(2)}}(x) = g_{\sigma^{(2)}}(c_2 x)$.

$\bar{S}_1$ and $\bar{S}_2$ satisfy the conditions claimed above, i.e. $\forall (x,y) \in \bar{S}_1 : |x| < 1/2$ and $\forall (x,y) \in \bar{S}_2 : |x| > 2$.

Let ϵ_i be the minimal distance of the point $(x_i, y_i) \in \bar{S}_1$ to some shattering polynomial in $\{\bar{f}_{\sigma^{(1)}}\}$, i.e.

$$\epsilon_i = \min_{f\in\{\bar{f}_{\sigma^{(1)}}\}} |f - y_i|.$$

Similarly let δ_j be defined for each point $(x_j, y_j) \in \bar{S}_2$ by

$$\delta_j = \min_{g\in\{\bar{g}_{\sigma^{(2)}}\}} |g - y_j|.$$

Again, by Remark 3 we can assume that $\epsilon_i, \delta_j > 0$.

Our goal is to unite the two sets so that their union is shattered by polynomials derived by adding two polynomials shattering the two single sets, i.e. the polynomial shattering the first set of the union may not interfere with the shattering of the second set and vice versa.

We define a polynomial $F(x)$ to be an upper bound on the shattering polynomials for $\bar{S}_1$ in the region according to $\bar{S}_2$, i.e.

$$F(x) > \max_{f(x)\in\{\bar{f}_{\sigma^{(1)}}\}} |f(x)| \quad \text{for all } |x| \geq 2,$$

in order to have an upper bound on the influence of the polynomials shattering $\bar{S}_1$ on the shattering of the set $\bar{S}_2$.

For some even integer N we transform the set $\bar{S}_2$ into the set $\bar{S}_2^N$ by

$$(x_j, y_j) \in \bar{S}_2 \quad \Longrightarrow \quad (x_j, x_j^N \cdot y_j) \in \bar{S}_2^N.$$

Since N is even, the set $\bar{S}_2^N$ is shattered by the set of t_2-sparse polynomials $\{x^N \cdot \bar{g}_{\sigma^{(2)}}\}$ and the minimal distance of the point $(x_j, y_j) \in \bar{S}_2^N$ to some shattering polynomial in $\{x^N \bar{g}_{\sigma^{(2)}}\}$ is $x_j^N \cdot \delta_j$.

We choose the parameter N to be large enough so that the following two conditions are satisfied:

- The polynomials $\{x^N \cdot \bar{g}_{\sigma^{(2)}}\}$ may not interfere with the shattering of $\bar{S}_1$, i.e

  $$x_i^N \cdot \bar{g}_{\sigma^{(2)}} < \epsilon_i \quad \text{ for all } (x_i, y_i) \in \bar{S}_1 \text{ and for all } \sigma^{(2)} \in \{+,-\}^{d_2},$$

  i.e. let G be the maximum of the absolute values of the polynomials $\{\bar{g}_{\sigma^{(2)}}(x)\}$ for $|x| \leq 1/2$ and ϵ the maximum over all ϵ_i. Then we choose N according to

  $$G \cdot (\frac{1}{2})^N < \epsilon, \qquad \text{i.e.} \quad N > \log_2(\frac{G}{\epsilon}).$$

- The polynomials $\{\bar{f}_{\sigma^{(1)}}\}$ may not interfere with the shattering of $\bar{S}_2^N$, i.e

  $$F(x_j) < x_j^N \cdot \delta_j \quad \text{ for all } (x_j, y_j) \in \bar{S}_2^N.$$

  We can choose such an N since all the x_j's have absolute value greater than 1 and N is even.

Let $\bar{S}_1 = \{(x_i', y_i')\}_{i=1,\ldots,d_1}$ with $x_1' < \ldots < x_{d_1}'$ and let $\bar{S}_2^N = \{(x_j'', y_j'')\}_{j=1,\ldots,d_2}$ with $x_1'' < \ldots < x_{d_2}''$.

Let $S = \bar{S}_1 \cup \bar{S}_2^N$. $S = \{(x_k, y_k)\}_{k=1,\ldots,d_1+d_2}$ and $x_1 < \ldots < x_{d_1+d_2}$. For $\sigma \in \{+,-\}^{d_1+d_2}$ we define $\sigma_1 \in \{+,-\}^{d_1}$ and $\sigma_2 \in \{+,-\}^{d_2}$ by

$$\sigma_1(i) = \sigma(k) \text{ iff } x_k = x_i' \qquad \text{and} \qquad \sigma_2(j) = \sigma(k) \text{ iff } x_k = x_j''.$$

S is of size d_1+d_2 and is shattered by the set of (t_1+t_2)-sparse polynomials $\{h_\sigma\}_{\sigma \in \{+,-\}^{d_1+d_2}}$, where $h_\sigma = f_{\sigma_1} + x^N g_{\sigma_2}$. Hence the $\mathrm{VC}_{\geq}$ dimension of the class of (t_1+t_2)-sparse polynomials is at least $\mathrm{VC}_{\geq}(\mathcal{P}_{t_1}) + \mathrm{VC}_{\geq}(\mathcal{P}_{t_2})$. □

We are now able to state our lower bound on the $\mathrm{VC}_{\geq}$ dimension of t-sparse polynomials.

Lemma 5. *The $\mathrm{VC}_{\geq}$ dimension of t-sparse polynomials is at least $3t$.*

Proof. By Lemma 4, we have $\mathrm{VC}_{\geq}(\mathcal{P}_t) \geq t \cdot \mathrm{VC}_{\geq}(\mathcal{P}_1)$. With Lemma 2 we have $\mathrm{VC}_{\geq}(\mathcal{P}_t) \geq 3t$. □

2.2 Upper bounds on $\mathbf{VC}_{\geq}(\mathcal{P}_t)$

In order to develop upper bounds on the $\mathrm{VC}_{\geq}$ dimension of t-sparse polynomials we make intensive use of upper bounds on the number of zeros of t-sparse polynomials.

Let $f_t = \sum_{i=1}^{t} c_i x^{e_i}$ be a t-sparse polynomial over the real numbers. f_t is said to be *axis symmetric* (or *even*) iff $f_t(x) = f_t(-x)$ (i.e. $\forall i = 1, \ldots, t:\ e_i$ is even) and f_t is said to be *point symmetric* (or *odd*) iff $f_t(x) = -f_t(-x)$ (i.e. $\forall i = 1, \ldots, t:\ e_i$ is odd).

Lemma 6. Let f_t be a t-sparse polynomial over the real numbers. Let $N(f_t)$ denote the number of real zeros of f_t.

(1) $N(f_t) \leq 2t - 1$. There are t-sparse polynomials with $2t - 1$ zeros.

(2) Let $N(f_t) = 2t - 1$. Then f_t is either odd or even.

(3) Let $N(f_t) = 2t - 1$. Then f_t changes sign on each of its zeros iff f_t is odd, and changes sign at each of its zeros except at the origin iff f_t is even.

(4) If $f_t = c$ has $2t$ solutions, c a constant, then f_t is even.

Proof. We show the statements by induction on t.

The case $t = 1$ is trivial. So assume Lemma 6 holds for all $k \leq t - 1$.

Let $f_t(x) = \sum_{i=1}^{t} c_i x^{e_i}$ with $e_1 < e_2 < \ldots < e_t$. Rewrite

$$f_t(x) = (\underbrace{\sum_{i=2}^{t} c_i x^{e_i - e_1} + c_1}_{g_{t-1}(x)}) \cdot x^{e_1}.$$

g_{t-1} is $(t-1)$-sparse, so is its derivation g'_{t-1}. g'_{t-1} has at most $2t - 3$ real zeros, therefore $g_{t-1} + c_1$ has at most $2t - 2$ zeros. Hence the number of zeros of f_t is bounded by $2t - 1$. On the other hand there are polynomials g_{t-1} s.t. $g_{t-1} + c_1$ has exactly $2t - 2$ zeros. Hence there are t-sparse polynomials with exactly $2t - 1$ real zeros. This proves the first statement.

Assume $N(f_t) = 2t - 1$. Then $(g_{t-1}(x) + c_1)$ has to have $2t - 2$ zeros, hence by (4) $(g_{t-1}(x) + c_1)$ is even. This proves (2).

Since $(g_{t-1}(x) + c_1)$ is even, $(g_{t-1}(x) + c_1)$ changes sign at each of its zeros except at the origin. Hence f_t changes sign at each of its zeros iff e_1 is odd, and changes sign at each of its zeros except at the origin iff e_1 is even. This proves (3).

If $f_t = c$ has $2t$ solutions then f_t' has $2t-1$ zeros and has to change sign on each of its zeros. Therefore f_t' is odd and f_t is even. This proves (4). □

We now state the upper bound on the $\mathrm{VC}_{\geq}$ dimension of t-sparse polynomials over the real numbers.

Lemma 7. *The $\mathrm{VC}_{\geq}$ dimension of t-sparse polynomials is at most $4t-1$.*

Proof. First, we prove that $\mathrm{VC}_{\geq}(\mathcal{P}_t) \leq 4t$.

Let $d = \mathrm{VC}_{\geq}(\mathcal{P}_t)$ and $S = \{(x_i, y_i)\}_{i=1,\ldots,d}$, where $x_1 < x_2 < \ldots < x_d$ be a set of points of size d shattered by t-sparse polynomials. Let f_1 and f_2 be t-sparse polynomials satisfying the two alternating labelings of the points in S. Then between each pair of points (x_i, y_i) and (x_{i+1}, y_{i+1}) in S, f_1 and f_2 must intersect. That is, $(f_1 - f_2)$ has at least $d-1$ zeros. Since $(f_1 - f_2)$ is $2t$-sparse, the number of zeros is bounded by $4t-1$ by Lemma 6. That is, $d \leq 4t$, proving that $\mathrm{VC}_{\geq}(\mathcal{P}_t) \leq 4t$.

Now, we show that there is no set of $4t$ points shattered by t-sparse polynomials.

Assume that $\mathrm{VC}_{\geq}(\mathcal{P}_t) = 4t$. Let $S = \{(x_i, y_i)\}_{i=1,\ldots,4t}$ where $x_1 < x_2 < \ldots < x_{4t}$ be a set of points shattered by t-sparse polynomials. Let f_1 and f_2 be t-sparse polynomials satisfying the two alternating labelings $\sigma_1 = (+,-,+,-,\cdots,+,-)$ and $\sigma_2 = (-,+,-,+,\cdots,-,+)$. f_1 and f_2 have $4t-1$ intersection points at $z_1 < z_2 < \ldots < z_{4t-1}$, and $(f_1 - f_2)$ has to change its sign at each of its zeros. Hence, by Lemma 6 (3), $(f_1 - f_2)$ is odd, therefore both f_1 and f_2 are point symmetric.

We define regions R_i to be the set of points between two neighboring intersection points and bounded by f_1 and f_2, i.e. (let $z_0 = -\infty$, $z_{4t} = +\infty$)

$$R_i := \{\ (x, y) \mid x \in [z_{i-1}, z_i],\ y \in [\min(f_1(x), f_2(x)), \max(f_1(x), f_2(x))]\ \}, \quad i = 1, \ldots, 4t.$$

We note that $(x_i, y_i) \in R_i$. R_i is point-symmetric to R_{4t-i} ($R_i = \{(-x,-y) | (x,y) \in R_{4t-i}\}$).

Consider the labelings γ_1 and γ_2 on S:

$$\gamma_1 = (\underbrace{-,+,-,+,\ldots,-,+}_{2t}, \underbrace{+,-,+,-,\ldots,+,-}_{2t}),$$

$$\gamma_2 = \gamma_1^{-1} = (\underbrace{+,-,+,-,\ldots,+,-}_{2t}, \underbrace{-,+,-,+,\ldots,-,+}_{2t}).$$

We prove that there are no t-sparse polynomials satisfying γ_1 and γ_2 on S. For purpose of contradiction assume that g_1, g_2 are t-sparse polynomials satisfying γ_1 and γ_2 on S.

g_1 and g_2 have at least $4t-2$ intersection points. Let c_1, c_2 be the constant terms of g_1, g_2, $G = g_1 - g_2 - (c_1 - c_2)$ and $c = c_2 - c_1$.

Assume $c_1 \neq 0$ and $c_2 \neq 0$. Then G is $(2t-2)$-sparse. Hence by Lemma 6, $G = c$ holds for at most $4t-4$ points, contradicting the assertion that g_1 and g_2 have at least $4t-2$ intersection points.

Assume $c_1 = 0$ and $c_2 \neq 0$. Then G is $(2t-1)$-sparse, and $G = c$ has to hold for at least $4t-2$ points. Hence by Lemma 6 G is even, so both g_1 and g_2 are axis symmetric. Let $v_1 < v_2 < \ldots < v_{2t-1} < v_{2t+1} < \ldots < v_{4t-1}$ be the intersection points of g_1 and g_2. We define regions T_i, similar to the definition of the regions R_i, to be the set of points between two neighboring intersection points of g_1 and g_2 bounded by g_1 and g_2, i.e. (let $v_0 = -\infty$, $v_{2t} = 0$, $v_{4t} = +\infty$)

$$T_j := \{\, (x,y) \mid x \in [v_{j-1}, v_j],\ y \in [\min(g_1(x), g_2(x)), \max(g_1(x), g_2(x))]\,\}, \quad j = 1, \ldots, 4t.$$

(We divided the region between v_{2t-1} and v_{2t+1} into two regions because of numbering reasons.) T_j is axis symmetric to T_{4t-j} ($T_j = \{(-x,y) | (x,y) \in T_{4t-j}\}$).

We note that $(x_i, y_i) \in T_i$. Since $(x_i, y_i) \in R_i$, the intersection $R_i \cup T_i \neq \emptyset$ for $i = 1, \ldots, 4t$. Suppose $R_i \cap \{y = 0\} = \emptyset$. Then $T_i \cap \{y = 0\} \neq \emptyset$ since the set R_i is point symmetric to R_{4t-i} and the region T_i is axis symmetric to T_{4t-i}.

For each $1 \leq i \leq 4t$, $R_i \cap \{y = 0\} \neq \emptyset$ forces f_1 or f_2 to have at least two zeros in $[z_{i-1}, z_i]$. Therefore by Lemma 6, there are at least $2t+2$ regions R_i such that $R_i \cap \{y = 0\} = \emptyset$. By a similar counting argument there are at most $2t-2$ regions such that $T_i \cap \{y = 0\} \neq \emptyset$ contradicting that for each i with $R_i \cap \{y = 0\} = \emptyset$ we have $T_i \cap \{y = 0\} \neq \emptyset$. Hence we contradicted the assumption $c_1 = 0$ and $c_2 \neq 0$. The case $c_1 \neq 0$ and $c_2 = 0$ is symmetric to this case.

Now assume both $c_1 = 0$ and $c_2 = 0$. For $i = 1, \ldots, 4t$ we define

$$R_i^+ := \{\, (x,y) \mid (x,y) \in R_i,\ y \geq g_1(x)\,\} \quad \text{and} \quad R_i^- := \{\, (x,y) \mid (x,y) \in R_i,\ y < g_1(x)\,\}.$$

Since g_1 is point symmetric to the origin, R_i^+ is point symmetric to R_{4t-i}^-. We show that $R_i^+, R_i^- \neq \emptyset$. Assume $R_i^+ = \emptyset$. Then $\gamma_1(i) = -$. Furthermore $R_{4t-i}^- = \emptyset$ and $\gamma_1(4t-i) = +$ contradicting the construction of γ_1.

We define for g_2 the sets $\bar{R}_i^+$ and $\bar{R}_i^-$ similar to R_i^+ and R_i^-. As above, $\bar{R}_i^+, \bar{R}_i^- \neq \emptyset$. Analogously, since $\gamma_2 = \gamma_1^{-1}$, the intersections $R_i^* \cap \bar{R}_i^*$ may not be empty. Hence g_1 and g_2 have to intersect in each region R_i, i.e. $(g_1 - g_2)$ has to have at least $4t$ zeros which contradicts the assumption that g_1 and g_2 are t-sparse.

Therefore there are no t-sparse polynomials satisfying γ_1 and γ_2 on S. Hence there is no set of $4t$ points shattered by t-sparse polynomials. $\square$

We note that the bounds derived in this section remain valid when restricted to t-sparse polynomials over the rational numbers and t-sparse polynomials over the integers.

Let $\mathcal{P}_t^+$ denote the class of t-sparse polynomials over the positive real numbers. Then, following the proofs in this section we can derive the exact $\mathrm{VC}_{\geq}$ dimension of the class $\mathcal{P}_t^+$:

$$\mathrm{VC}_{\geq}(\mathcal{P}_t^+) = 2t.$$

Furthermore our results can be transferred to sparse rational functions. Let $\mathcal{R}_t$ denote the class of real rational functions with t-sparse numerator and t-sparse denominator. Following the proof of Lemma 7, we derive the upper bound

$$\mathrm{VC}_{\geq}(\mathcal{R}_t) \leq 4t^2.$$

Because of the finiteness of the $\mathrm{VC}_{\geq}$ dimension of the classes $\mathcal{P}_t$ and $\mathcal{R}_t$, we can state the following theorem without explicitly giving learning algorithms for these classes ([BEHW 87]):

Theorem 8. *The classes of sparse polynomials and sparse rational functions are uniformly learnable.*

3 Approximating Polynomial Regression

One of the central problems in computational regression theory is the problem of determining the number of terms in an arranged system of functions. The most important case of this problem is the approximation of polynomial regression (cf. [Vap 82], pp. 254–258). For underlying definitions and terminology see [Po 84], [Vap 82].

The classical scheme of approximating polynomial regression, which involves the determination of the true degree n of regression and the expansion in a system of n orthogonal polynomials of degree $1, 2, \ldots, n$, can be successfully implemented only when large samples are used. The reason for this is the (possibly) large degree of regression and therefore the large ϵ-VC dimension (capacity) of the class of polynomials of degree n. The problem for small samples remained open. Here we are giving solution to this problem.

Hence this problem reduces to the determination of the ϵ-VC dimension (capacity) of sparse polynomials (independent of the degree). We prove linear bounds on the ϵ-VC dimension of t-sparse polynomials, and, as a direct consequence, derive the surprising result that the regression function can be approximated uniformly by sparse polynomials.

First, we bound the ϵ-VC dimension of the class $\mathcal{P}_t$. As a remark, we note that ϵ-$\mathrm{VC}(\mathcal{P}_t)$ is independent of ϵ, i.e.

$$\epsilon_1\text{-}\mathrm{VC}(\mathcal{P}_t) = \epsilon_2\text{-}\mathrm{VC}(\mathcal{P}_t), \qquad \epsilon_1, \epsilon_2 > 0.$$

PROOF. Let $d = \epsilon_1$-VC$(\mathcal{P}_t)$ and $S_{\epsilon_1} = \{(x_i, y_i)\}_{i=1,\ldots,d}$ be a set of points shattered (in ϵ_1-sense) by the set of t-sparse polynomials $\{f_\sigma\}_{\sigma\in\{+,-\}^d} \subset \mathcal{P}_t$, i.e.

$$\forall i = 1,\ldots,d \ \forall\sigma \in \{+,-\}^d \ : \ |f_\sigma(x_i) - y_i| \begin{cases} \leq \epsilon_1 & \sigma(i) = + \\ > \epsilon_1 & \sigma(i) = - \end{cases} .$$

Hence

$$\forall i = 1,\ldots,d \ \forall\sigma \in \{+,-\}^d \ : \ |\frac{\epsilon_1}{\epsilon_2} f_\sigma(x_i) - \frac{\epsilon_1}{\epsilon_2} y_i| \begin{cases} \leq \epsilon_2 & \sigma(i) = + \\ > \epsilon_2 & \sigma(i) = - \end{cases} ,$$

therefore the set $S_{\epsilon_2} = \{(x, \frac{\epsilon_1}{\epsilon_2} y) \mid (x, y) \in S_{\epsilon_1}\}$ of size d is shattered (in ϵ_2-sense) by the set of t-sparse polynomials $\{\frac{\epsilon_1}{\epsilon_2} f_\sigma\}_{\sigma\in\{+,-\}^d} \subset \mathcal{P}_t^{\mathcal{P}}$. Hence ϵ_2-VC$(\mathcal{P}_t) \geq \epsilon_1$-VC$(\mathcal{P}_t)$ and vice versa.
□

Lemma 9.

$$\epsilon\text{-VC}(\mathcal{P}_t) \geq \text{VC}_{\geq}(\mathcal{P}_t).$$

PROOF. Let $d = \text{VC}_{\geq}(\mathcal{P}_t)$ and $S_{\geq} = \{(x_i, y_i)\}_{i=1,\ldots,d}$ be a set of points shattered (in $\geq$-sense) by the set of t-sparse polynomials $\{f_\sigma\}_{\sigma\in\{+,-\}^d} \subset \mathcal{P}_t$, i.e.

$$\forall i = 1,\ldots,d \ \forall\sigma \in \{+,-\}^d \ : \ f_\sigma(x_i) - y_i \begin{cases} \leq 0 & \sigma(i) = + \\ > 0 & \sigma(i) = - \end{cases} .$$

Let ϵ be defined by

$$\epsilon = \max_{i=1,\ldots,d} \ \max_{\sigma,\sigma(i)=+} \ y_i - f_\sigma(x_i).$$

Then

$$\forall i = 1,\ldots,d \ \forall\sigma \in \{+,-\}^d \ : \ |f_\sigma(x_i) - (y_i - \epsilon)| \begin{cases} \leq \epsilon & \sigma(i) = + \\ > \epsilon & \sigma(i) = - \end{cases} .$$

Hence the set $S_\epsilon = \{(x, y - \epsilon) \mid (x, y) \in S_{\geq}\}$ of size d is shattered (in ϵ-sense) by the set of t-sparse polynomials $\{f_\sigma\}_{\sigma\in\{+,-\}^d} \subset \mathcal{P}_t$. Hence ϵ-VC$(\mathcal{P}_t) \geq \text{VC}_{\geq}(\mathcal{P}_t)$. □

We introduce the following lemma to derive an upper bound on ϵ-VC$(\mathcal{P}_t)$.

Lemma 10. Let $S = \{(x_i, y_i)\}_{i=1,\ldots,4}$ where $x_1 < x_2 < x_3 < x_4$ be a set of points and let $\sigma_1 = (+,-,-,+)$, $\sigma_2 = (-,+,+,-)$, $\sigma_3 = (+,-,+,-)$, $\sigma_4 = (-,+,-,+)$ be labelings on S. Let $\{f_i\}_{i=1,\ldots,4}$ be continuous functions satisfying σ_i on S (in ϵ-sense). Then there is an intersection of either (f_1, f_2) or (f_1, f_3) or (f_1, f_4) or (f_3, f_4) in the interval (x_1, x_4).

PROOF. Consider the 2^8 cases for $f_i(x_j) > y_j + \epsilon$ or $f_i(x_j) < y_j - \epsilon$ if $\sigma_i(j) = -$. □

Lemma 11. *The ϵ-VC dimension of t-sparse polynomials is at most $48t - 9$.*

PROOF. Let $d = \epsilon$-VC$(\mathcal{P}_t)$ and $S = \{(x_i, y_)\}_{i=1,\ldots,d}$ where $x_1 < x_2 < \ldots < x_d$ be a set of points of size d shattered by t-sparse polynomials. Consider the labelings $\sigma_1 =$

$(+,-,-,+,-,-,+,-,-,\ldots)$, $\sigma_2 = (-,+,+,-,+,+,-,+,+,\ldots)$, $\sigma_3 = (+,-,+,-,+,-,\ldots)$ and $\sigma_4 = (-,+,-,+,-,+,\ldots)$. Let $f_1,\ldots,f_4$ be t-sparse polynomials satisfying $\sigma_1,\ldots,\sigma_4$. Then, by Lemma 10 there are two polynomials with at least $d/12$ intersections and with Lemma 6 we conclude that $d \le 12(4t-1)+3 = 48t-9$. □

Now we state our main theorem, solving the problem of Vapnik [Vap 82].

Theorem 12. *The polynomial regression can be approximated uniformly by sparse polynomials for small samples.* □

Acknowledgements

We thank Manuel Blum, Allan Borodin, Sally Floyd, Les Valiant and Manfred Warmuth for the number of interesting conversations. The discussion with Vladimir Vapnik has led us to the solution of the problem of the polynomial regression function.

References

[BT 88] Ben-Or, M., Tiwari, P., *A Deterministic Algorithm for Sparse Multivariate Polynomial Interpolation*, Proc. 20^{th} ACM STOC (1988), pp. 301–309.

[BEHW 87] Blumer, A., Ehrenfeucht, A., Hausler, D., Warmuth, M., *Learnability and the Vapnik-Chervonenkis Dimension*, UC Santa Cruz, Tech. Rep. UCSC-CRL-87-20, 1987.

[BT 89] Borodin, A., Tiwari, P., *On the Decidability of Sparse Univariate Polynomial Interpolation*, IBM Research Report RC 14923 (#66763), Sep. 1989.

[Fl 89] Floyd, S., *Space-Bounded Learning and the Vapnik-Chervonenkis Dimension*, Manuscript, International Computer Science Institute, Berkeley, 1989.

[GK 87] Grigoriev, D.Yu., Karpinski, M., *The Matching Problem for Bipartite Graphs with Polynomially Bounded Permannets Is in NC*, Proc. 28^{th} IEEE FOCS (1987), pp. 166–172.

[GKS 88] Grigoriev, D.Yu., Karpinski, M., Singer, M., *Fast Parallel Algorithms for Sparse Multivariate Polynomial Interpolation over Finite Fields*, University of Bonn, Research Report No. 8523-CS, 1988.

[Ha 89] Hausler, D., *Generalizing the PAC Model: Sample Size Bounds From Metric Dimension-based Uniform Convergence Results*, Proc. 30^{th} IEEE FOCS (1989), pp. 40–45.

[KV 89] Kearns, M., Valiant, L.G., *Cryptographic Limitations on Learning Boolean Formulae and Finite Automata*, Proc. 21^{st} ACM STOC (1989), pp. 433–444.

[Po 84] Pollard, D., *Convergence of Stochastic Processes*, Springer-Verlag, 1984.

[Val 84] Valiant, L.G., *A Theory of the Learnable*, Comm. ACM, 27(11), 1984, pp. 1134–1142.

[Vap 82] Vapnik, V.N., *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, 1982.

[Vap 89] Vapnik, V.N., *Inductive Principles of the Search for Empirical Dependences (Methods Based on Weak Convergence of Probability Measures)*, Proc. of the 2^{nd} Workshop on Computational Learning Theory, 1989.

[VC 71] Vapnik, V.N., Chervonenkis, A.Y., *On the Uniform Convergence of Relative Frequencies of Events and their Probabilities*, Th. Prob. and its Appl., 16(2), 1971, pp. 264–280.