Descartes' Rule of Signs for Radial Basis Function Neural Networks

Michael Schmitt

Lehrstuhl Mathematik und Informatik, Fakultät für Mathematik
Ruhr-Universität Bochum, D–44780 Bochum, Germany
http://www.ruhr-uni-bochum.de/lmi/mschmitt/
mschmitt@lmi.ruhr-uni-bochum.de

Abstract

We establish versions of Descartes' rule of signs for radial basis function (RBF) neural networks. The RBF rules of signs provide tight bounds for the number of zeros of univariate networks with certain parameter restrictions. Moreover, they can be used to infer that the Vapnik-Chervonenkis (VC) dimension and pseudo-dimension of these networks are no more than linear. This contrasts with previous work showing that RBF neural networks with two and more input nodes have superlinear VC dimension. The rules give rise also to lower bounds for network sizes, thus demonstrating the relevance of network parameters for the complexity of computing with RBF neural networks.

1 Introduction

First published in 1637, Descartes' rule of signs is an astoundingly simple and yet powerful method to estimate the number of zeros of a univariate polynomial (Descartes, 1954; Struik, 1986). Precisely, the rule says that the number of positive zeros of a real univariate polynomial is equal to the number of sign changes in the sequence of its coefficients, or is less than this number by a multiple of two (see, e.g., Henrici, 1974).

We show that versions of Descartes' rule of signs can be established for radial basis function (RBF) neural networks. We focus on networks with standard, that is, Gaussian hidden units with parameters satisfying certain conditions. In particular, we formulate RBF rules of signs for networks with uniform widths. In the strongest version, they yield that for every univariate function computed by these networks, the number of zeros is bounded by the number of sign changes occurring in a sequence of the output weights. A similar rule is stated for networks with uniform centers. We further show that the bounds given by these rules are tight, a fact also known for the original Descartes' rule (Anderson et al., 1998; Grabiner, 1999). The RBF rules of signs are presented in Section 2.

We employ the rules to derive bounds on the Vapnik-Chervonenkis (VC) dimension and the pseudo-dimension of RBF neural networks. These dimensions measure the diversity of a function class and are basic concepts studied in theories of learning and generalization (see, e.g. Anthony and Bartlett, 1999). The pseudo-dimension has also applications in approximation theory (Maiorov and Ratsaby, 1999; Dung, 2001). We show that the VC dimension and the pseudo-dimension of RBF neural networks with a single input node and uniform widths is no more than linear in the number of parameters. The same holds for networks with uniform centers. These results nicely contrast with previous work showing that these dimensions grow superlinearly for RBF neural networks with more than one input node and at least two different width values (Schmitt, 2002). In particular, the bound established there is $\Omega(W \log k)$, where W is the number of parameters and k the number of hidden units. Further, for the networks considered here, the linear upper bound improves a result of Bartlett and Williamson (1996) who have shown that Gaussian RBF neural networks restricted to discrete inputs from $\{-D, \ldots, D\}$ have these dimensions bounded by $O(W \log(WD))$. Moreover, the linear bounds are tight and considerably smaller than the bound $O(W^2 k^2)$ known for unrestricted Gaussian RBF neural networks (Karpinski and Macintyre, 1997). We establish the linear bounds in Section 3.

The computational power of neural networks is a well recognized and frequently studied phenomenon. Not so well understood, however, is the question how large a network must be to perform a specific task. The rules of signs and the VC dimension bounds provided here give partly rise to answers in terms of lower bounds on the size of RBF networks. These bounds cast light on the relevance of network parameters for the complexity of computing with networks having uniform widths. In particular, we show that a network with zero bias must be more than twice as large to be as powerful as a network with variable bias. Further, to have the capabilities of a network with nonuniform widths, it must be larger in size by a factor of at least 1.5. The latter result is of particular significance since, as shown by Park and Sandberg (1991), the universal approximation capabilities of RBF networks hold even for networks with uniform widths (see also Park and Sandberg, 1993). The bounds for the complexity of computing are derived in Section 4.

The results presented here are concerned with networks having a single input node. Networks of this type are not, and have never been, of minor importance. This is evident from the numerous simulation experiments performed to assess the capabilities of neural networks for approximating univariate functions. For instance, Broomhead and Lowe (1988), Moody and Darken (1989), Mel and

Omohundro (1991), Wang and Zhu (2000), and Li and Leiss (2001) have done such studies using RBF networks. Moreover, the best known lower bound for the VC dimension of neural networks, being quadratic in the number of parameters, holds, among others, for univariate networks (Koiran and Sontag, 1997). Thus, the RBF rules of signs provide new insights for a neural network class that is indeed relevant in theory and practice.

2 Rules of Signs for RBF Neural Networks

We consider a type of RBF neural network known as Gaussian or standard RBF neural network.

Definition 1. *A radial basis function (RBF) neural network computes functions over the reals of the form*

$$w_0 + w_1 \exp\left(-\frac{\|x - c_1\|^2}{\sigma_1^2}\right) + \cdots + w_k \exp\left(-\frac{\|x - c_k\|^2}{\sigma_k^2}\right),$$

where $\|\cdot\|$ denotes the Euclidean norm. Each exponential term corresponds to a hidden unit with center c_i and width σ_i, respectively. The input nodes are represented by x, and $w_0, \ldots, w_k$ denote the output weights of the network. Centers, widths, and ouptut weights are the network parameters. The network has uniform widths if $\sigma_1 = \cdots = \sigma_k$; it has uniform centers if $c_1 = \cdots = c_k$. The parameter w_0 is also referred to as the bias or threshold of the network.

Note that the output weights and the widths are scalar parameters, whereas x and the centers are tuples if the network has more than one input node. In the following, all functions have a single input variable unless indicated otherwise.

Definition 2. *Given a finite sequence $w_1, \ldots, w_k$ of real numbers, a change of sign occurs at position j if there is some $i < j$ such that the conditions*

$$w_i w_j \;<\; 0$$

and

$$w_l \;=\; 0, \quad \text{for } i < l < j,$$

are satisfied.

Thus we can refer to the number of changes of sign of a sequence as a well defined quantity.

2.1 Uniform Widths

We first focus on networks with zero bias and show that the number of zeros of a function computed by an RBF network with uniform widths does not exceed the number of changes of sign occurring in a sequence of its output weights. Moreover, the difference of the two quantities is always even. Thus, we obtain for RBF networks a precise rephrasing of Descartes' rule of signs for polynomials (see, e.g., Henrici, 1974, Theorem 6.2d). We call it the strong RBF rule of signs. In contrast to Descartes' rule, the RBF rule is not confined to the strictly positive numbers, but valid for whole real domain. Its derivation generalizes Laguerre's proof of Descartes' rule (Pólya and Szegő, 1976, Part V, Chapter 1, No. 77).

Theorem 1 (Strong RBF Rule of Signs for Uniform Widths). *Consider a function $f : \mathbb{R} \to \mathbb{R}$ computed by an RBF neural network with zero bias and k hidden units having uniform width. Assume that the centers be ordered such that $c_1 \leq \cdots \leq c_k$ and let $w_1, \ldots, w_k$ denote the associated output weights. Let s be the number of changes of sign of the sequence $w_1, \ldots, w_k$ and z the number of zeros of f. Then $s - z$ is even and nonnegative.*

Proof. Without loss of generality, let k be the smallest number of RBF units required for computing f. Then we have $w_i \neq 0$ for $1 \leq i \leq k$, and due to the equality of the widths, $c_1 < \cdots < c_k$. We first show by induction on s that $z \leq s$ holds.

Obviously, if $s = 0$ then f has no zeros. Let $s > 0$ and assume the statement is valid for $s - 1$ changes of sign. Suppose that there is a change of sign at position i. As a consequence of the assumptions above, $w_{i-1}w_i < 0$ holds and there is some real number b satisfying $c_{i-1} < b < c_i$. Clearly, f has the same number of zeros as the function g defined by

$$g(x) \quad = \quad \exp\left(\frac{x^2 - 2bx}{\sigma^2}\right) \cdot f(x),$$

where σ is the common width of the hidden units in the network computing f. According to Rolle's theorem, the derivative g' of g must have at least $z - 1$ zeros. Then, the function h defined as

$$h(x) \quad = \quad \exp\left(-\frac{x^2 - 2bx}{\sigma^2}\right) \cdot g'(x)$$

has at least $z - 1$ zeros as well. We can write this function as

$$h(x) \quad = \quad \sum_{\lambda=1}^{k} w_\lambda \frac{2(c_\lambda - b)}{\sigma^2} \exp\left(-\frac{(x - c_\lambda)^2}{\sigma^2}\right).$$

Hence, h is computed by an RBF network with zero bias and k hidden units having width σ and centers $c_1, \ldots, c_k$. Since b satisfies $c_{i-1} < b < c_i$ and the sequence $w_1, \ldots, w_k$ has a change of sign at position i, there is no change of sign at position i in the sequence of the output weights

$$w_1 \frac{2(c_1 - b)}{\sigma^2}, \ldots, w_{i-1} \frac{2(c_{i-1} - b)}{\sigma^2}, w_i \frac{2(c_i - b)}{\sigma^2}, \ldots, w_k \frac{2(c_k - b)}{\sigma^2}.$$

At positions different from i, there is a change of sign if and only if the sequence $w_1, \ldots, w_k$ has a change of sign at this position. Hence, the number of changes of sign in the output weights for h is equal to $s-1$. By the induction hypothesis, h has at most $s-1$ zeros. As we have argued above, h has at least $z-1$ zeros. This yields $z \leq s$, completing the induction step.

It remains to show that $s - z$ is even. For $x \to +\infty$, the term

$$w_k \exp\left(-\frac{(x - c_k)^2}{\sigma^2}\right)$$

becomes in absolute value larger than the sum of the remaining terms so that its sign determines the sign of f. Similarly, for $x \to -\infty$ the behavior of f is governed by the term

$$w_1 \exp\left(-\frac{(x - c_1)^2}{\sigma^2}\right).$$

Thus, the number of zeros of f is even if and only if both terms have the same sign. And this holds if and only if the number of sign changes in the sequence $w_1, \ldots, w_k$ is even. This proves the statement. □

The above result establishes the number of sign changes as upper bound for the number of zeros. It is easy to see that this bound cannot be improved. If the width is sufficiently small or the intervals between the centers are sufficiently large, the sign of the output of the network at some center c_i is equal to the sign of w_i. Thus we can enforce a zero between any pair of adjacent centers c_{i-1}, c_i through $w_{i-1} w_i < 0$.

Corollary 2. *The parameters of every univariate RBF neural network can be chosen such that all widths are the same and the number of changes of sign in the sequence of the output weights $w_1, \ldots, w_k$, ordered according to $c_1 \leq \cdots \leq c_k$, is equal to the number of zeros of the function computed by the network.*

Next, we consider networks with freely selectable bias. The following result is the main step in deriving a bound on the number of zeros for these networks. It deals with a more general network class where the bias can be an arbitrary polynomial.

Theorem 3. *Suppose $p : \mathbb{R} \to \mathbb{R}$ is a polynomial of degree l and $g : \mathbb{R} \to \mathbb{R}$ is computed by an RBF neural network with zero bias and k hidden units having uniform width. Then the function $p + g$ has at most $l + 2k$ zeros.*

Proof. We perform induction on k. For the sake of formal simplicity, we say that the all-zero function is the (only) function computed by the network with zero hidden units. Then for $k = 0$ the function $p + g$ is a polynomial of degree l that, by the fundamental theorem of algebra, has no more than l zeros.

Now, let g be computed by an RBF network with $k > 0$ hidden units, so that $p + g$ can be written as

$$p + \sum_{i=1}^{k} w_i \exp\left(-\frac{(x-c_i)^2}{\sigma^2}\right).$$

Let z denote the number of zeros of this function. Multiplying with $\exp((x - c_k)^2/\sigma^2)$ we obtain the function

$$\exp\left(\frac{(x-c_k)^2}{\sigma^2}\right) p + w_k + \sum_{i=1}^{k-1} w_i \exp\left(-\frac{(x-c_i)^2 - (x-c_k)^2}{\sigma^2}\right)$$

that must have z zeros as well. Differentiating this function with respect to x yields

$$\exp\left(\frac{(x-c_k)^2}{\sigma^2}\right)\left(\frac{2(x-c_k)}{\sigma^2} p + p'\right) + \sum_{i=1}^{k-1} w_i \frac{2(c_i-c_k)}{\sigma^2} \exp\left(-\frac{(x-c_i)^2 - (x-c_k)^2}{\sigma^2}\right),$$

which, according to Rolle's theorem, must have at least $z - 1$ zeros. Again, multiplication with $\exp(-(x - c_k)^2/\sigma^2)$ leaves the number of zeros unchanged and we get with

$$\frac{2(x-c_k)}{\sigma^2} p + p' + \sum_{i=1}^{k-1} w_i \frac{2(c_i-c_k)}{\sigma^2} \exp\left(-\frac{(x-c_i)^2}{\sigma^2}\right)$$

a function of the form $q + h$, where q is a polynomial of degree $l + 1$ and h is computed by an RBF network with zero bias and $k - 1$ hidden units having uniform width. By the induction hypothesis, $q + h$ has at most $l + 1 + 2(k - 1) = l + 2k - 1$ zeros. Since $z - 1$ is a lower bound, it follows that $z \leq l + 2k$ as claimed. $\square$

From this we immediately have a bound for the number of zeros for RBF networks with nonzero bias. We call this fact the weak RBF rule of signs. In contrast to the strong rule, the bound is not in terms of the number of sign changes but in terms of the number of hidden units. The naming, however, is justified since the proofs for both rules are very similar. As for the strong rule, the bound of the weak rule cannot be improved.

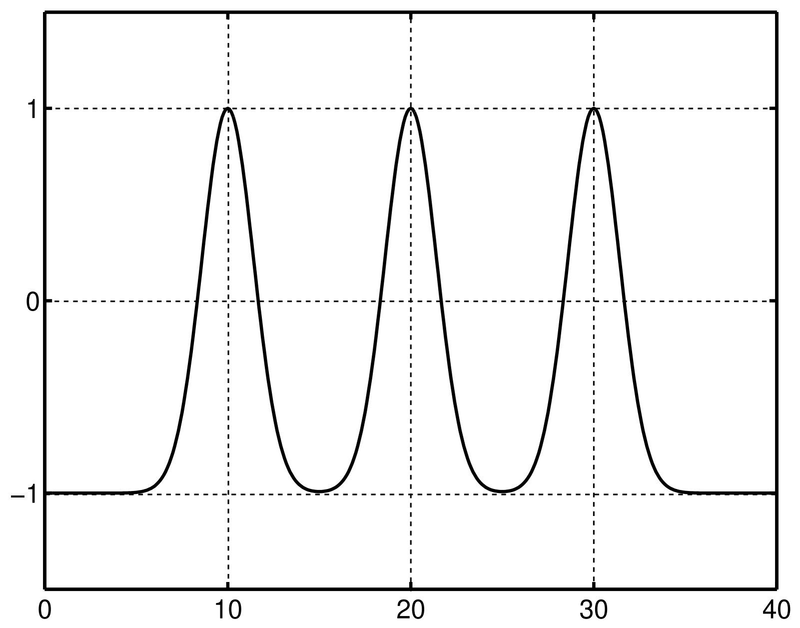

Figure 1: Optimality of the weak RBF rule of signs for uniform widths: A function having twice as many zeros as hidden units.

Corollary 4 (Weak RBF Rule of Signs for Uniform Widths). *Let function $f : \mathbb{R} \to \mathbb{R}$ be computed by an RBF neural network with arbitrary bias and k hidden units having uniform width. Then f has at most $2k$ zeros. Furthermore, this bound is tight.*

Proof. The upper bound follows from Theorem 3 since $l = 0$. The lower bound is obtained using a network that has a negative bias and all other output weights positive. For instance, -1 for the bias and 2 for the weights are suitable. An example for $k = 3$ is displayed in Figure 1. For sufficiently small widths and large intervals between the centers, the output value is close to 1 for inputs close to a center and approaches -1 between the centers and toward $+\infty$ and $-\infty$. Thus, each hidden unit gives rise to two zeros. $\square$

The function in Figure 1 has only one change of sign: The bias is negative, all other weights are positive. Consequently, we realize that a strong rule of signs does not hold for networks with nonzero bias.

2.2 Uniform Centers

We now establish a rule of signs for networks with uniform centers. Here, in contrast to the rules for uniform widths, the bias plays a role in determining the number of sign changes.

Theorem 5 (RBF Rule of Signs for Uniform Centers). *Suppose $f : \mathbb{R} \to \mathbb{R}$ is computed by an RBF neural network with arbitrary bias and k hidden units having uniform centers. Assume the ordering $\sigma_1^2 \leq \cdots \leq \sigma_k^2$ and let s denote the number of sign changes of the sequence $w_0, w_1, \ldots, w_k$. Then f has at most $2s$ zeros.*

Proof. Given f as supposed, consider the function g obtained by introducing the new variable $y = (x - c)^2$, where c is the common center. Then

$$g(y) \quad = \quad w_0 + w_1 \exp(-\sigma_1^{-2} y) + \cdots + w_k \exp(-\sigma_k^{-2} y).$$

Let z be the number of zeros of g and s the number of sign changes of the sequence $w_0, w_1, \ldots, w_k$. Similarly as in Theorem 1 we deduce $z \leq s$ by induction on s. Assume there is a change of sign at position i and let b satisfy $\sigma_{i-1}^{-2} < b < \sigma_i^{-2}$. (Without loss of generality, the widths are all different.) By Rolle's theorem, the derivative of the function $\exp(by) \cdot g(y)$ has at least $z - 1$ zeros. Multiplying this derivative with $\exp(-by)$ yields

$$w_0 b + w_1 (b - \sigma_1^{-2}) \exp(-\sigma_1^{-2} y) + \cdots + w_k (b - \sigma_k^{-2}) \exp(-\sigma_k^{-2} y)$$

with at least $z - 1$ zeros. Moreover, this function has at most $s - 1$ sign changes. (Note that $b > 0$.) This completes the induction.

Now, each zero $y \neq 0$ of g gives rise to exactly two zeros $x = c \pm \sqrt{y}$ for f. (By definition, $y \geq 0$.) Further, $g(0) = 0$ if and only if $f(c) = 0$. Thus, f has at most $2s$ zeros. □

That the bound is optimal can be seen from Figure 2 showing a function having three hidden units with center 20, widths $\sigma_1 = 1, \sigma_2 = 4, \sigma_3 = 8$, and output weights $w_0 = 1/2, w_1 = -3/4, w_2 = 2, w_3 = -2$. The example can easily be generalized to any number of hidden units.

3 Linear Bounds for VC Dimension and Pseudo-Dimension

We apply the RBF rules of signs to obtain bounds for the VC dimension and the pseudo-dimension that are linear in the number of hidden units. First, we give the definitions.

A set S is said to be shattered by a class $\mathcal{F}$ of $\{0,1\}$-valued functions if for every dichotomy (S_0, S_1) of S (where $S_0 \cup S_1 = S$ and $S_0 \cap S_1 = \emptyset$) there is some $f \in \mathcal{F}$ such that $f(S_0) \subseteq \{0\}$ and $f(S_1) \subseteq \{1\}$. Let $\mathrm{sgn} : \mathbb{R} \to \{0,1\}$ denote the function satisfying $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = 0$ otherwise.

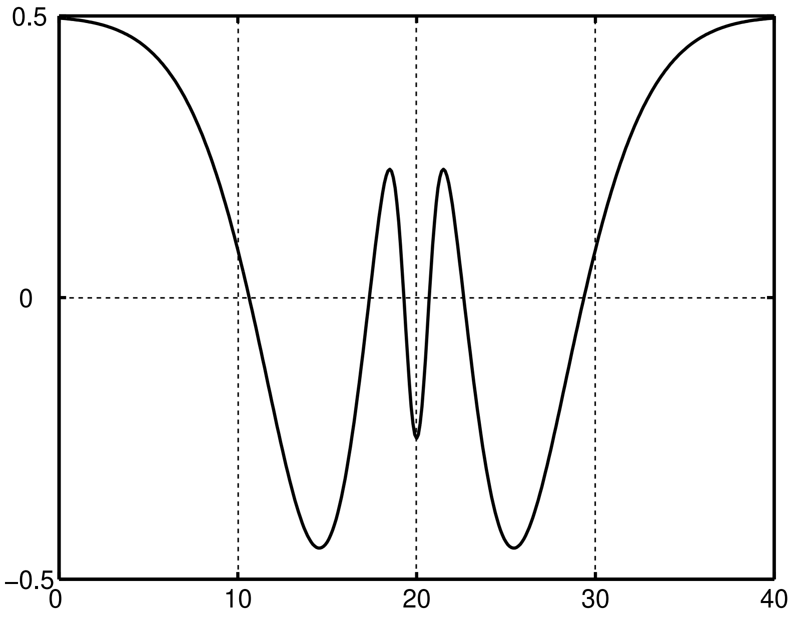

Figure 2: Optimality of the RBF rule of signs for uniform centers: A function having twice as many zeros as sign changes.

Definition 3. *Given a class $\mathcal{G}$ of real-valued functions, the VC dimension of $\mathcal{G}$ is the largest integer m such that there exists a set S of cardinality m that is shattered by the class $\{\operatorname{sgn} \circ g : g \in \mathcal{G}\}$. The pseudo-dimension of $\mathcal{G}$ is the VC dimension of the class $\{(x, y) \mapsto g(x) - y : g \in \mathcal{G}\}$.*

We extend this definition to neural networks in the obvious way: The VC dimension of a neural network is defined as the VC dimension of the class of functions computed by the network (obtained by assigning all possible values to its parameters); analogously for the pseudo-dimension. It is evident that the VC dimension of a neural network is not larger than its pseudo-dimension.

3.1 Uniform Widths

The RBF rules of signs give rise to the following VC dimension bounds.

Theorem 6. *The VC dimension of every univariate RBF neural network with k hidden units, variable but uniform widths, and zero bias does not exceed k. For variable bias, the VC dimension is at most $2k + 1$.*

Proof. Let $S = \{x_1, \ldots, x_m\}$ be shattered by an RBF neural network with k hidden units. Assume that $x_1 < \cdots < x_m$ and consider the dichotomy (S_0, S_1)

defined as

$$x_i \in S_0 \iff i \text{ even}.$$

Let f be the function of the network that implements this dichotomy. We want to avoid that $f(x_i) = 0$ for some i. This is the case only if $x_i \in S_1$. Since $f(x_i) < 0$ for all $x_i \in S_0$, we can slightly adjust some output weight such that the resulting function g of the network still induces the dichotomy (S_0, S_1) but does not yield zero on any element of S. Clearly, g must have a zero between x_i and x_{i+1} for every $1 \le i \le m-1$. If the bias is zero then, by Theorem 1, g has at most $k-1$ zeros. This implies $m \le k$. From Corollary 4 we have $m \le 2k+1$ for nonzero bias. □

As a consequence of this result, we now know the VC dimension of univariate uniform-width RBF neural networks exactly. In particular, we observe that the use of a bias more than doubles their VC dimension.

Corollary 7. *The VC dimension of every univariate RBF neural network with k hidden units and variable but uniform widths equals k if the bias is zero, and $2k+1$ if the bias is variable.*

Proof. Theorem 6 has shown that k and $2k+1$ are upper bounds, respectively. The lower bound k for zero bias is implied by Corollary 2. For nonzero bias, consider a set S with $2k+1$ elements and a dichotomy (S_0, S_1). If $|S_1| \le k$, we put a hidden unit with positive output weight over each element of S_1 and let the bias be negative. Figure 1, for instance, implements the dichotomy $(\{5, 15, 25, 35\}, \{10, 20, 30\})$. If $|S_0| \le k$, we proceed the other way round with negative output weights and a positive bias. □

It is easy to see that the result holds also if the widths are fixed. This observation is helpful in the following where we address the pseudo-dimension and establish linear bounds for networks with uniform and fixed widths.

Theorem 8. *Consider a univariate RBF neural network with k hidden units having the same fixed width. For zero bias, the pseudo-dimension d of the network satisfies $k \le d \le 2k$; if the bias is variable then $2k+1 \le d \le 4k+1$.*

Proof. The lower bounds follow from the fact that the VC dimension is a lower bound for the pseudo-dimension and that Corollary 7 is also valid for fixed widths. For the upper bound we use an idea that has been previously employed by Karpinski and Werther (1993) and Andrianov (1999) to obtain pseudo-dimension bounds in terms of zeros of functions. Let $\mathcal{F}_{k,\sigma}$ denote the class of functions computed by the network with k hidden units and width σ. Assume the set $S = \{(x_1, y_1), \ldots, (x_m, y_m)\} \subseteq \mathbb{R}^2$ is shattered by the class

$$\{(x, y) \mapsto \operatorname{sgn}(f(x) - y) : f \in \mathcal{F}_{k,\sigma}\}.$$

Let $x_1 < \cdots < x_m$ and consider the dichotomy (S_0, S_1) defined as

$$(x_i, y_i) \in S_0 \iff i \text{ even}.$$

Since S is shattered, there exist functions $f_1, f_2 \in \mathcal{F}_{k,\sigma}$ that implement the dichotomies (S_0, S_1) and (S_1, S_0), respectively. That is, we have for $i = 1, \dots, m$

$$f_2(x_i) \geq y_i > f_1(x_i) \iff i \text{ even},$$
$$f_1(x_i) \geq y_i > f_2(x_i) \iff i \text{ odd}.$$

This implies

$$(f_1(x_i) - f_2(x_i)) \cdot (f_1(x_{i+1}) - f_2(x_{i+1})) \quad < \quad 0$$

for $i = 1, \dots, m-1$. Therefore, $f_1 - f_2$ must have at least $m-1$ zeros. On the other hand, since $f_1, f_2 \in \mathcal{F}_{k,\sigma}$, the function $f_1 - f_2$ can be computed by an RBF network with $2k$ hidden units having equal width. Considering networks with zero bias, it follows from the strong RBF rule of signs (Theorem 1) that $f_1 - f_2$ has at most $2k-1$ zeros. For variable bias, the weak rule of signs (Corollary 4) bounds the number of zeros by $4k$. This yields $m \leq 2k$ for zero and $m \leq 4k+1$ for variable bias. □

3.2 Uniform Centers

Finally, we state the linear VC dimension and pseudo-dimension bounds for networks with uniform centers.

Corollary 9. *The VC dimension of every univariate RBF neural network with k hidden units and variable but uniform centers is at most $2k+1$. For networks with uniform and fixed centers, the pseudo-dimension is at most $4k+1$.*

Proof. Using the RBF rule of signs for uniform centers (Theorem 5), the VC dimension bound is derived following the proof of Theorem 6, the bound for the pseudo-dimension is obtained as in Theorem 8. □

Employing techniques from Corollary 7, it is not hard to derive that k is a lower bound for the VC dimension of these networks and, hence, also for the pseudo-dimension.

4 Computational Complexity

In this section we demonstrate how the above results can be used to derive lower bounds on the size of networks required for simulating other networks that have more parameters. We focus on networks with uniform widths. Lower bounds for networks with uniform centers can be obtained analogously.

We know from Corollary 7 that introducing a variable bias increases the VC dimension of uniform-width RBF networks by a factor of more than two. As an immediate consequence, we have a lower bound on the size of networks with zero bias for simulating networks with nonzero bias. It is evident that a network with zero bias cannot simulate a network with nonzero bias on the entire real domain. The question makes sense, however, if the inputs are restricted to a compact subset, as done in statements about uniform approximation capabilities of neural networks.

Corollary 10. *Every univariate RBF neural network with variable bias and k hidden units requires at least $2k+1$ hidden units to be simulated by a network with zero bias and uniform widths.*

In the following we show that the number of zeros of functions computed by networks with zero bias and nonuniform widths need not be bounded by the number of sign changes. Thus, the strong RBF rule of signs does not hold for these networks.

Lemma 11. *For every $k \geq 1$ there is a function $f_k : \mathbb{R} \to \mathbb{R}$ that has at least $3k-1$ zeros and can be computed by an RBF neural network with zero bias, $2k$ hidden units and two different width values.*

Proof. Consider the function

$$f_k(x) \quad = \quad \sum_{i=1}^{k}(-1)^i \left(w_1 \exp\left(\frac{(x-c_i)^2}{\sigma_1^2}\right) - w_2 \exp\left(\frac{(x-c_i)^2}{\sigma_2^2}\right)\right).$$

Clearly, it is computed by a network with zero bias and $2k$ hidden units having widths σ_1 and σ_2. It can also immediately be seen that the parameters can be instantiated such that f_k has at least $3k-1$ zeros. An example for $k=2$ is shown in Figure 3. Here, the centers are $c_1 = 5, c_2 = 15$, the widths are $\sigma_1 = 1/2, \sigma_2 = 2$, and the output weights are $w_1 = 2, w_2 = 1$. By juxtaposing copies of this figure, examples for larger values of k are easily obtained. □

For zero bias, the family of functions constructed in the previous proof gives rise to a lower bound on the size of networks with uniform widths simulating networks with nonuniform widths. In particular, uniform-width networks must be at least a factor of 1.5 larger to be as powerful as nonuniform-width networks. This separates the computational capabilities of the two network models and demonstrates the power of the width parameter.

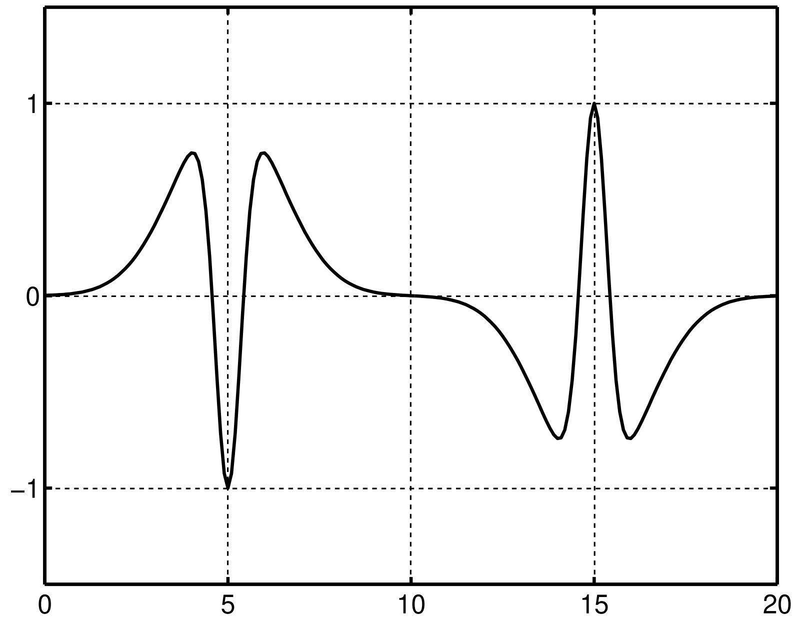

Figure 3: Function f_2 of Lemma 11 has five zeros and is computed by an RBF network with four hidden units using two different widths.

Corollary 12. *A univariate zero-bias RBF neural network with $2k$ hidden units of arbitrary width cannot be simulated by a zero-bias network with uniform widths and less than $3k$ hidden units. This holds even if the nonuniform-width network uses only two different width values.*

Proof. Consider again the functions f_k defined in the proof of Lemma 11. Each f_k is computed by a network with $2k$ hidden units using two different width values and has $3k-1$ zeros. By virtue of the strong RBF rule of signs (Theorem 1), at least $3k$ hidden units are required for a network with uniform widths to have $3k-1$ zeros. □

The lower bounds derived here are concerned with zero-bias networks only. We remark at this point that it remains open whether and to what extent nonuniform widths increase the computational power of networks with nonzero bias.

5 Conclusion

We have derived rules of signs for various types of univariate RBF neural networks. These rules have been shown to yield tight bounds for the number of zeros of functions computed by these networks. Two quantities have turned out to be crucial: On the one hand, the number of sign changes in a sequence of the output weights for networks with zero bias and uniform widths and for networks with uniform centers; on the other hand, the number of hidden units for networks with nonzero bias and uniform widths. It remains as the most challenging open problem to find a rule of signs for networks with nonuniform widths and nonuniform centers.

Using the rules of signs, linear bounds on the VC dimension and pseudo-dimension of the studied networks have been established. The smallest bound has been found for networks with zero bias and uniform widths, for which the VC dimension is equal to the number of hidden units and, hence, less than half the number of parameters. Further, introducing one more parameter, the bias, more than doubles this VC dimension. The pseudo-dimension bounds have been obtained for networks with fixed width or fixed centers only. Since the pseudo-dimension is defined via the VC dimension using one more variable, getting bounds for more general networks might be a matter of looking at higher input dimensions.

Finally, we have calculated lower bounds for sizes of networks simulating other networks with more parameters. These results assume that the simulations are performed exactly. In view of the approximation capabilities of neural networks it is desirable to have such bounds also for notions of approximative computation.

Acknowledgment

This work has been supported in part by the ESPRIT Working Group in Neural and Computational Learning II, NeuroCOLT2, No. 27150.

References

Anderson, B., Jackson, J., and Sitharam, M. (1998). Descartes' rule of signs revisited. *American Mathematical Monthly*, 105:447–451.

Andrianov, A. (1999). On pseudo-dimension of certain sets of functions. *East Journal on Approximations*, 5:393–402.

Anthony, M. and Bartlett, P. L. (1999). *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, Cambridge.

Bartlett, P. L. and Williamson, R. C. (1996). The VC dimension and pseudodimension of two-layer neural networks with discrete inputs. *Neural Computation*, 8:625–628.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355.

Descartes, R. (1954). *The Geometry of René Descartes with a Facsimile of the First Edition*. Dover Publications, New York, NY. Translated by D. E. Smith and M. L. Latham.

Dung, D. (2001). Non-linear approximations using sets of finite cardinality or finite pseudo-dimension. *Journal of Complexity*, 17:467–492.

Grabiner, D. J. (1999). Descartes' rule of signs: Another construction. *American Mathematical Monthly*, 106:854–856.

Henrici, P. (1974). *Applied and Computational Complex Analysis 1: Power Series, Integration, Conformal Mapping, Location of Zeros*. Wiley, New York, NY.

Karpinski, M. and Macintyre, A. (1997). Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks. *Journal of Computer and System Sciences*, 54:169–176.

Karpinski, M. and Werther, T. (1993). VC dimension and uniform learnability of sparse polynomials and rational functions. *SIAM Journal on Computing*, 22:1276–1285.

Koiran, P. and Sontag, E. D. (1997). Neural networks with quadratic VC dimension. *Journal of Computer and System Sciences*, 54:190–198.

Li, S.-T. and Leiss, E. L. (2001). On noise-immune RBF networks. In Howlett, R. J. and Jain, L. C., editors, *Radial Basis Function Networks 1: Recent Developments in Theory and Applications*, volume 66 of *Studies in Fuzziness and Soft Computing*, chapter 5, pages 95–124. Springer-Verlag, Berlin.

Maiorov, V. and Ratsaby, J. (1999). On the degree of approximation by manifolds of finite pseudo-dimension. *Constructive Approximation*, 15:291–300.

Mel, B. W. and Omohundro, S. M. (1991). How receptive field parameters affect neural learning. In Lippmann, R. P., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 3*, pages 757–763. Morgan Kaufmann, San Mateo, CA.

Moody, J. and Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1:281–294.

Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3:246–257.

Park, J. and Sandberg, I. W. (1993). Approximation and radial-basis-function networks. *Neural Computation*, 5:305–316.

Pólya, G. and Szegő, G. (1976). *Problems and Theorems in Analysis II: Theory of Functions. Zeros. Polynomials. Determinants. Number Theory. Geometry.* Springer-Verlag, Berlin.

Schmitt, M. (2002). Neural networks with local receptive fields and superlinear VC dimension. *Neural Computation*, 14:919–956.

Struik, D. J., editor (1986). *A Source Book in Mathematics, 1200–1800.* Princeton University Press, Princeton, NJ.

Wang, Z. and Zhu, T. (2000). An efficient learning algorithm for improving generalization performance of radial basis function neural networks. *Neural Networks*, 13:545–553.