Radial Basis Function Neural Networks Have Superlinear VC Dimension

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 superlinear lower bounds on the Vapnik-Chervonenkis (VC) dimension of neural networks with one hidden layer and local receptive field neurons. As the main result we show that every reasonably sized standard network of radial basis function (RBF) neurons has VC dimension $\Omega(W \log k)$, where W is the number of parameters and k the number of nodes. This significantly improves the previously known linear bound. We also derive superlinear lower bounds for networks of discrete and continuous variants of center-surround neurons. The constants in all bounds are larger than those obtained thus far for sigmoidal neural networks with constant depth.

The results have several implications with regard to the computational power and learning capabilities of neural networks with local receptive fields. In particular, they imply that the pseudo dimension and the fat-shattering dimension of these networks is superlinear as well, and they yield lower bounds even when the input dimension is fixed. The methods developed here appear suitable for obtaining similar results for other kernel-based function classes.

1 Introduction

Although there exists already a large collection of Vapnik-Chervonenkis (VC) dimension bounds for neural networks, it has not been known thus far whether the VC dimension of radial basis function (RBF) neural networks is superlinear. Major reasons for this might be that previous results establishing superlinear bounds are based on methods geared to sigmoidal neurons or consider networks having an unrestricted number of layers [3, 10, 13, 22]. RBF neural networks, however, differ from other neural network types in two characteristic features (see, e.g., Bishop [4], Ripley [21]): There is only one hidden layer and the neurons have local receptive fields. In particular, the neurons are not of the sigmoidal type (see Koiran and Sontag [10] for a rather general definition of a sigmoidal activation function that does not capture radial basis functions).

Beside sigmoidal networks, RBF networks are among the major neural network types used in practice. They are appreciated because of their impressive capabilities in function approximation and learning that have been well studied in theory and practice (see, e.g., [5, 9, 15, 16, 18, 19, 20]). Sigmoidal neural networks are known having VC dimension that is superlinear in the number of network parameters, even when there is only one hidden layer [22]. Since the VC dimension of single neurons is linear, this superlinearity witnesses the enormous computational capabilities that emerge when neurons cooperate in networks. The VC dimension of RBF networks has been studied earlier by Anthony and Holden [2], Lee et al. [11, 12], and Erlich et al. [7]. In particular, Erlich et al. [7] established a linear lower bound, and Anthony and Holden [2] posed as an open problem whether a superlinear bound can be shown.

In this paper we prove that the VC dimension of RBF networks is indeed superlinear. Precisely, we show that every network with n input nodes, W parameters, and one hidden layer of k RBF neurons, where $k \leq 2^{(n+2)/2}$, has VC dimension at least $(W/12)\log(k/8)$. Thus, the cooperative network effect enhancing the computational power of sigmoidal networks is now confirmed for RBF networks, too. Furthermore, the result has consequences for the complexity of learning with RBF networks, all the more since it entails the same lower bound for the pseudo dimension and the fat-shattering dimension. (See Anthony and Bartlett [1] for implications of VC dimension bounds, and the relationship between the VC dimension, the pseudo dimension, and the fat-shattering dimension.)

We do not derive the result for RBF networks directly but take a major detour. We first consider networks consisting of a different type of locally processing units, the so-called binary center-surround receptive field (CSRF) neurons. These are discrete models of neurons found in the visual system of mammals (see, e.g., Nicholls et al. [17, Chapter 16], Tessier-Lavigne [23]). In Section 3 we establish a superlinear VC dimension bound for CSRF neural networks showing that every network having W parameters and k hidden nodes has VC dimension at least $(W/5)\log(k/4)$, where $k \leq 2^{n/2}$. Then in Section 4 we look at a continuous variant of the CSRF neuron known as the difference-of-Gaussians (DOG) neuron, which computes the weighted difference of two concentric Gaussians. This type of unit is widely used as a continuous model of neurons in the visual pathway (see, e.g., Glezer [8], Marr [14]). Utilizing the result for CSRF networks we show that DOG networks have VC dimension at least $(W/5)\log(k/4)$ as well. Finally, the above claimed lower bound for RBF networks is then immediately obtained.

We note that regarding the constants, the bounds for CSRF and DOG networks are larger than for RBF networks. Further, all bounds we derive for networks of local receptive field neurons have larger constant factors than those known for sigmoidal networks of constant depth thus far. For comparison, sigmoidal networks are known that have one hidden layer and VC dimension at least $(W/32)\log(k/4)$; for two hidden layers a VC dimension of at least $(W/132)\log(k/16)$ has been found (see Anthony and Bartlett [1, Section 6.3]). Finally, the results obtained here give rise to linear lower bounds for local receptive field neural networks when the input dimension is fixed.

2 Definitions and Notation

Let $||u||$ denote the Euclidean norm of vector u. A *Gaussian radial basis function (RBF) neuron* computes a function $g_{\mathrm{RBF}} : \mathbb{R}^{2n+1} \to \mathbb{R}$ defined as

$$g_{\mathrm{RBF}}(c, \sigma, x) \quad = \quad \exp\left(-\frac{\|x-c\|^2}{\sigma^2}\right),$$

with input variables $x_1, \ldots, x_n$, and parameters $c_1, \ldots, c_n$ (the *center*) and $\sigma > 0$ (the *width*). A *difference-of-Gaussians (DOG) neuron* is defined as a function $g_{\mathrm{DOG}} : \mathbb{R}^{2n+4} \to \mathbb{R}$ computed by the weighted difference of two RBF neurons with equal centers, that is,

$$g_{\mathrm{DOG}}(c, \sigma, \tau, \alpha, \beta, x) \quad = \quad \alpha g_{\mathrm{RBF}}(c, \sigma, x) - \beta g_{\mathrm{RBF}}(c, \tau, x).$$

A *binary center-surround receptive field (CSRF) neuron* computes a function $g_{\mathrm{CSRF}} : \mathbb{R}^{2n+2} \to \{0, 1\}$ defined as

$$g_{\mathrm{CSRF}}(c, a, b, x) = 1 \quad \Longleftrightarrow \quad a \leq \|x - c\| \leq b$$

with *center* $(c_1, \ldots, c_n)$, *center radius* $a > 0$, and *surround radius* $b > a$. We also refer to it as *off-center on-surround neuron* and call for given parameters c, a, b the set $\{x : g_{\mathrm{CSRF}}(c, a, b, x) = 1\}$ the *surround region* of the neuron.

We consider *neural networks with one hidden layer* computing functions of the form $f : \mathbb{R}^{W+n} \to \mathbb{R}$, where W is the number of network parameters, n the number of input nodes, and

$$f(w, y, x) \quad = \quad w_0 + w_1 h_1(y, x) + \cdots + w_k h_k(y, x).$$

The k hidden nodes compute functions $h_1, \ldots, h_k \in \{g_{\mathrm{RBF}}, g_{\mathrm{DOG}}, g_{\mathrm{CSRF}}\}$. (Each hidden node "selects" its parameters from y, which comprises all parameters of the hidden nodes.) Note that if $h_i = g_{\mathrm{RBF}}$ for $i = 1, \ldots, k$, this is the standard form of a radial basis function neural network. The network has a linear output node with parameters $w_0, \ldots, w_k$ also known as the *output weights*. For simplicity we sometimes refer to all network parameters as weights.

An $(n-1)$-dimensional *hyperplane* in $\mathbb{R}^n$ is given by a vector $(w_0, \ldots, w_n) \in \mathbb{R}^{n+1}$ and defined as the set

$$\{x \in \mathbb{R}^n : w_0 + w_1 x_1 + \cdots + w_n x_n = 0\}.$$

An $(n-1)$-dimensional *hypersphere* in $\mathbb{R}^n$ is represented by a center $c \in \mathbb{R}^n$ and a radius $r > 0$, and defined as the set

$$\{x \in \mathbb{R}^n : \|x - c\| = r\}.$$

We also consider hyperplanes and hyperspheres in $\mathbb{R}^n$ with a dimension $k < n-1$. In this case, a k-dimensional hyperplane (hypersphere) is the intersection of two $(k+1)$-dimensional hyperplanes (hyperspheres), assuming that the intersection is non-empty. (For hyperspheres we additionally require that the intersection is not a single point.)

A *dichotomy* of a set $S \subseteq \mathbb{R}^n$ is a pair (S_0, S_1) of subsets such that $S_0 \cap S_1 = \emptyset$ and $S_0 \cup S_1 = S$. A class $\mathcal{F}$ of functions mapping $\mathbb{R}^n$ to $\{0, 1\}$ *shatters* S if every dichotomy (S_0, S_1) of S is *induced* by some $f \in \mathcal{F}$ (i.e., satisfying $f(S_0) \subseteq \{0\}$ and $f(S_1) \subseteq \{1\}$). The *Vapnik-Chervonenkis (VC) dimension* of $\mathcal{F}$ is the cardinality of the largest set shattered by $\mathcal{F}$. The VC dimension of a neural network $\mathcal{N}$ is defined as the VC dimension of the class of functions computed by $\mathcal{N}$, where the output is made binary using some threshold.

We use "ln" to denote the natural logarithm and "log" for the logarithm to base 2.

3 Lower Bounds for CSRF Networks

In this section we consider one-hidden-layer networks of binary center-surround receptive field neurons. The main result requires a property of certain finite sets of points.

Definition 1. *A set S of m points in $\mathbb{R}^n$ is said to be* in spherically general position *if the following two conditions are satisfied:*

(1) For every $k \leq \min(n, m-1)$ and every $(k+1)$-element subset $P \subseteq S$, there is no $(k-1)$-dimensional hyperplane containing all points in P.

(2) For every $l \leq \min(n, m-2)$ and every $(l+2)$-element subset $Q \subseteq S$, there is no $(l-1)$-dimensional hypersphere containing all points in Q.

Sets satisfying condition (1) are commonly referred to as being "in general position" (see, e.g., Cover [6]). For the VC dimension bounds we require sufficiently large sets in spherically general position. It is not hard to show that such sets exist.

Proposition 1. *For every $n, m \geq 1$ there exists a set $S \subseteq \mathbb{R}^n$ of m points in spherically general position.*

Proof. We perform induction on m. Clearly, every single point trivially satisfies conditions (1) and (2). Assume that some set $S \subseteq \mathbb{R}^n$ of cardinality m has been constructed. Then by the induction hypothesis, for every $k \leq \min(n, m)$, every k-element subset $P \subseteq S$ does not lie on a hyperplane of dimension less than $k-1$. Hence, every $P \subseteq S$, $|P| = k \leq \min(n, m)$, uniquely specifies a $(k-1)$-dimensional hyperplane H_P that includes P. The induction hypothesis implies further that no point in $S \setminus P$ lies on H_P. Analogously, for every $l \leq \min(n, m-1)$, every $(l+1)$-element subset $Q \subseteq S$ does not lie on a hypersphere of dimension less than $l-1$. Thus, every $Q \subseteq S$, $|Q| = l + 1 \leq \min(n, m-1) + 1$, uniquely determines an $(l-1)$-dimensional hypersphere B_Q containing all points in Q and none of the points in $S \setminus Q$.

To obtain a set of cardinality $m+1$ in spherically general position we observe that the union of all hyperplanes and hyperspheres considered above, that is, the union of all H_P and all B_Q for all subsets P and Q, has Lebesgue measure 0. Hence, there is some point $s \in \mathbb{R}^n$ not contained in any hyperplane H_P and not contained in any hypersphere B_Q. By adding s to S we then obtain a set of cardinality $m+1$ in spherically general position. $\square$

The following theorem establishes the major step for the superlinear lower bound.

Theorem 2. *Let $h, q, m \geq 1$ be arbitrary natural numbers. Suppose $\mathcal{N}$ is a network with one hidden layer consisting of binary CSRF neurons, where the number of hidden nodes is $h + 2^q$ and the number of input nodes is $m + q$. Assume further that the output node is linear. Then there exists a set of cardinality $hq(m+1)$ shattered by $\mathcal{N}$. This even holds if the output weights of $\mathcal{N}$ are fixed to 1.*

Proof. Before starting with the details we give a brief outline. The main idea is to imagine the set we want to shatter as being composed of groups of vectors, where the groups are distinguished by means of the first m components and the remaining q components identify the group members. We catch these groups by hyperspheres such that each hypersphere is responsible for up to $m+1$ groups. The condition of spherically general position will ensure that this works. The hyperspheres are then expanded to become surround regions of CSRF neurons. To induce a dichotomy of the given set, we split the groups. We do this for each group using the q last components in such a way that the points with designated output 1 stay within the surround region of the respective neuron and the points with designated output 0 are expelled from it. In order for this to succeed, we have to make sure that the displaced points do not fall into the surround region of some other neuron. The verification of the split operation will constitute the major part of the proof.

By means of Proposition 1 let $\{s_1, \ldots, s_{h(m+1)}\} \subseteq \mathbb{R}^m$ be in spherically general position. Let $e_1, \ldots, e_q$ denote the unit vectors in $\mathbb{R}^q$, that is, with a 1 in exactly one component and 0 elsewhere. We define the set S by

$$S = \{s_i : i = 1, \ldots, h(m+1)\} \times \{e_j : j = 1, \ldots, q\}.$$

Clearly, S is a subset of $\mathbb{R}^{m+q}$ and has cardinality $hq(m+1)$. To show that S is shattered by $\mathcal{N}$, let (S_0, S_1) be some arbitrary dichotomy of S. Consider an enumeration $M_1, \ldots, M_{2^q}$ of all subsets of the set $\{1, \ldots, q\}$. Let the function $f : \{1, \ldots, h(m+1)\} \to \{1, \ldots, 2^q\}$ be defined by

$$M_{f(i)} = \{j : s_i e_j \in S_1\},$$

where $s_i e_j$ denotes the vector resulting from the concatenation of s_i and e_j. We use f to define a partition of $\{s_1, \ldots, s_{h(m+1)}\}$ into sets T_k for $k = 1, \ldots, 2^q$ by

$$T_k = \{s_i : f(i) = k\}.$$

We further partition each set T_k into subsets $T_{k,p}$ for $p = 1, \ldots, \lceil |T_k|/(m+1) \rceil$, where each subset $T_{k,p}$ has cardinality $m+1$, except if $m+1$ does not divide $|T_k|$, in which case there is exactly one subset of cardinality less than $m+1$. Since there are at most $h(m+1)$ elements s_i, the partitioning of all T_k results in no more than h subsets of cardinality $m+1$. Further, the fact $k \le 2^q$ permits at most 2^q subsets of cardinality less than $m+1$. Thus, there are no more than $h + 2^q$ subsets $T_{k,p}$.

We employ one hidden node $H_{k,p}$ for each subset $T_{k,p}$. Thus the $h + 2^q$ hidden nodes of $\mathcal{N}$ suffice. Since $\{s_1, \ldots, s_{h(m+1)}\}$ is in spherically general position, there exists for each $T_{k,p}$ an $(m-1)$-dimensional hypersphere containing all points in $T_{k,p}$ and no other point. If $|T_{k,p}| = m+1$, this hypersphere is unique; if $|T_{k,p}| < m+1$, there is a unique $(|T_{k,p}| - 2)$-dimensional hypersphere which can be extended to an $(m-1)$-dimensional hypersphere that does not contain any further point. (Here we require condition (1) from the definition of spherically general position, otherwise no hypersphere of dimension $|T_{k,p}| - 2$ including all points of $T_{k,p}$ might exist.) Clearly, if $|T_{k,p}| = 1$, we can also extend this single point to an $(m-1)$-dimensional hypersphere not including any further point.

Suppose that $(c_{k,p}, r_{k,p})$ with center $c_{k,p}$ and radius $r_{k,p}$ represents the hypersphere associated with subset $T_{k,p}$. It is obvious from the construction above that all radii satisfy $r_{k,p} > 0$. Further, since the subsets $T_{k,p}$ are pairwise disjoint, there is some $\varepsilon > 0$ such that every point $s_i \in \{s_1, \ldots, s_{h(m+1)}\}$ and every just defined hypersphere $(c_{k,p}, r_{k,p})$ satisfy

$$\text{if } s_i \notin T_{k,p} \text{ then } |\, \|s_i - c_{k,p}\| - r_{k,p}| > \varepsilon. \qquad (1)$$

In other words, ε is smaller than the distance between any s_i and any hypersphere $(c_{k,p}, r_{k,p})$ that does not contain s_i. Without loss of generality we assume that ε

is sufficiently small such that

$$\varepsilon \quad \leq \quad \min_{k,p} r_{k,p}. \tag{2}$$

The parameters of the hidden nodes are adjusted as follows: We define the center $\widehat{c}_{k,p} = (\widehat{c}_{k,p,1}, \ldots, \widehat{c}_{k,p,m+q})$ of hidden node $H_{k,p}$ by assigning the vector $c_{k,p}$ to the first m components and specifying the remaining ones by

$$\widehat{c}_{k,p,m+j} \quad = \quad \begin{cases} 0 & \text{if} j \in M_k, \\ -\varepsilon^2/4 & \text{otherwise}, \end{cases}$$

for $j = 1, \ldots, q$. We further define new radii $\widehat{r}_{k,p}$ by

$$\widehat{r}_{k,p} \quad = \quad \sqrt{r_{k,p}^2 + (q - |M_k|)\left(\frac{\varepsilon}{2}\right)^4 + 1}$$

and choose some $\gamma > 0$ satisfying

$$\gamma \quad \leq \quad \min_{k,p} \frac{\varepsilon^2}{8\widehat{r}_{k,p}}. \tag{3}$$

The center and surround radii $\widehat{a}_{k,p}, \widehat{b}_{k,p}$ of the hidden nodes are then specified as

$$\begin{aligned} \widehat{a}_{k,p} \quad &= \quad \widehat{r}_{k,p} - \gamma, \\ \widehat{b}_{k,p} \quad &= \quad \widehat{r}_{k,p} + \gamma. \end{aligned}$$

Note that $\widehat{a}_{k,p} > 0$ holds, because $\varepsilon^2 < \widehat{r}_{k,p}^2$ implies $\gamma < \widehat{r}_{k,p}$.

This completes the assignment of parameters to the hidden nodes $H_{k,p}$. We now derive two inequalities concerning the relationship between ε and γ that we need in the following. First, we estimate $\varepsilon^2/2$ from below by

$$\begin{aligned} \frac{\varepsilon^2}{2} \quad &> \quad \frac{\varepsilon^2}{4} + \frac{\varepsilon^2}{64} \\ &> \quad \frac{\varepsilon^2}{4} + \frac{\varepsilon^4}{(8\widehat{r}_{k,p})^2} \quad \text{for all } k, p, \end{aligned}$$

where the last inequality is obtained from $\varepsilon^2 < \widehat{r}_{k,p}^2$. Using (3) for both terms on the right-hand side, we get

$$\frac{\varepsilon^2}{2} \quad > \quad 2\widehat{r}_{k,p}\gamma + \gamma^2 \quad \text{for all } k, p. \tag{4}$$

Second, from (2) we get

$$-r_{k,p}\varepsilon + \frac{\varepsilon^2}{2} \quad < \quad -\frac{\varepsilon^2}{4} \quad \text{for all } k, p,$$

and (3) yields

$$-\frac{\varepsilon^2}{4} \quad < \quad -2\widehat{r}_{k,p}\gamma \qquad \text{for all } k,p.$$

Putting the last two inequalities together and adding γ^2 to the right-hand side, we obtain

$$-r_{k,p}\varepsilon + \frac{\varepsilon^2}{2} \quad < \quad -2\widehat{r}_{k,p}\gamma + \gamma^2 \qquad \text{for all } k,p. \tag{5}$$

We next establish three facts about the hidden nodes.

Claim (i). *Let $s_i e_j$ be some point and $T_{k,p}$ some subset where $s_i \in T_{k,p}$ and $j \in M_k$. Then hidden node $H_{k,p}$ outputs 1 on $s_i e_j$.*

According to the definition of $\widehat{c}_{k,p}$, if $j \in M_k$, we have

$$\|s_i e_j - \widehat{c}_{k,p}\|^2 \quad = \quad \|s_i - c_{k,p}\|^2 + (q - |M_k|)\left(\frac{\varepsilon}{2}\right)^4 + 1.$$

The condition $s_i \in T_{k,p}$ implies $\|s_i - c_{k,p}\|^2 = r_{k,p}^2$, and thus

$$\begin{aligned} \|s_i e_j - \widehat{c}_{k,p}\|^2 \quad &= \quad r_{k,p}^2 + (q - |M_k|)\left(\frac{\varepsilon}{2}\right)^4 + 1 \\ &= \quad \widehat{r}_{k,p}^2. \end{aligned}$$

It follows that $\|s_i e_j - \widehat{c}_{k,p}\| = \widehat{r}_{k,p}$, and since $\widehat{a}_{k,p} < \widehat{r}_{k,p} < \widehat{b}_{k,p}$, point $s_i e_j$ lies within the surround region of node $H_{k,p}$. Hence, Claim (i) is shown.

Claim (ii). *Let $s_i e_j$ and $T_{k,p}$ satisfy $s_i \in T_{k,p}$ and $j \notin M_k$. Then hidden node $H_{k,p}$ outputs 0 on $s_i e_j$.*

From the assumptions we get here

$$\begin{aligned} \|s_i e_j - \widehat{c}_{k,p}\|^2 \quad &= \quad \|s_i - c_{k,p}\|^2 + (q - |M_k| - 1)\left(\frac{\varepsilon}{2}\right)^4 + \left(1 + \frac{\varepsilon^2}{4}\right)^2 \\ &= \quad r_{k,p}^2 + (q - |M_k|)\left(\frac{\varepsilon}{2}\right)^4 + 1 + \frac{\varepsilon^2}{2} \\ &= \quad \widehat{r}_{k,p}^2 + \frac{\varepsilon^2}{2}. \end{aligned}$$

Employing (4) on the right-hand side results in

$$\|s_i e_j - \widehat{c}_{k,p}\|^2 \quad > \quad \widehat{r}_{k,p}^2 + 2\widehat{r}_{k,p}\gamma + \gamma^2.$$

Hence, taking square roots we have $\|s_i e_j - \widehat{c}_{k,p}\| > \widehat{r}_{k,p} + \gamma$, implying that $s_i e_j$ lies outside the surround region of $H_{k,p}$. Thus, Claim (ii) follows.

Claim (iii). *Let $s_i e_j$ be some point and $T_{k,p}$ some subset such that $s_i \in T_{k,p}$. Then every hidden node $H_{k',p'}$ with $(k',p') \neq (k,p)$ outputs 0 on $s_i e_j$.*

Since $s_i \in T_{k,p}$ and s_i is not contained in any other subset $T_{k',p'}$, condition (1) implies

$$\|s_i - c_{k',p'}\|^2 > (r_{k',p'} + \varepsilon)^2 \quad \text{or} \quad \|s_i - c_{k',p'}\|^2 < (r_{k',p'} - \varepsilon)^2. \tag{6}$$

We distinguish between two cases: whether $j \in M_{k'}$ or not.

Case 1. If $j \in M_{k'}$ then by the definition of $\widehat{c}_{k',p'}$ we have

$$\|s_i e_j - \widehat{c}_{k',p'}\|^2 \;=\; \|s_i - c_{k',p'}\|^2 + (q - |M_{k'}|)\left(\frac{\varepsilon}{2}\right)^4 + 1.$$

From this, using (6) and the definition of $\widehat{r}_{k',p'}$ we obtain

$$\begin{array}{rcl} \|s_i e_j - \widehat{c}_{k',p'}\|^2 & > & \widehat{r}^2_{k',p'} + 2r_{k',p'}\varepsilon + \varepsilon^2 \\ & \text{or} & \\ \|s_i e_j - \widehat{c}_{k',p'}\|^2 & < & \widehat{r}^2_{k',p'} - 2r_{k',p'}\varepsilon + \varepsilon^2. \end{array} \tag{7}$$

We derive bounds for the right-hand sides of these inequalities as follows. From (4) we have

$$\varepsilon^2 \;>\; 4\widehat{r}_{k',p'}\gamma + 2\gamma^2,$$

which, after adding $2r_{k',p'}\varepsilon$ to the left-hand side and halving the right-hand side, gives

$$2r_{k',p'}\varepsilon + \varepsilon^2 \;>\; 2\widehat{r}_{k',p'}\gamma + \gamma^2. \tag{8}$$

From (2) we get $\varepsilon^2/2 < r_{k',p'}\varepsilon$, that is, the left-hand side of (5) is negative. Hence, we may double it to obtain from (5)

$$-2r_{k',p'}\varepsilon + \varepsilon^2 \;<\; -2\widehat{r}_{k',p'}\gamma + \gamma^2.$$

Using this and (8) in (7) leads to

$$\|s_i e_j - \widehat{c}_{k',p'}\|^2 > (\widehat{r}_{k',p'} + \gamma)^2 \quad \text{or} \quad \|s_i e_j - \widehat{c}_{k',p'}\|^2 < (\widehat{r}_{k',p'} - \gamma)^2.$$

And this is equivalent to

$$\|s_i e_j - \widehat{c}_{k',p'}\| > \widehat{b}_{k',p'} \quad \text{or} \quad \|s_i e_j - \widehat{c}_{k',p'}\| < \widehat{a}_{k',p'},$$

meaning that $H_{k',p'}$ outputs 0.

Case 2. If $j \notin M_{k'}$ then

$$\|s_i e_j - \widehat{c}_{k',p'}\|^2 \;=\; \|s_i - c_{k',p'}\|^2 + (q - |M_{k'}|)\left(\frac{\varepsilon}{2}\right)^4 + 1 + \frac{\varepsilon^2}{2}.$$

As a consequence of this together with (6) and the definition of $\widehat{r}_{k',p'}$ we get

$$\begin{array}{rcl} \|s_i e_j - \widehat{c}_{k',p'}\|^2 & > & \widehat{r}^2_{k',p'} + 2r_{k',p'}\varepsilon + \varepsilon^2 + \dfrac{\varepsilon^2}{2} \\ & \text{or} & \\ \|s_i e_j - \widehat{c}_{k',p'}\|^2 & < & \widehat{r}^2_{k',p'} - 2r_{k',p'}\varepsilon + \varepsilon^2 + \dfrac{\varepsilon^2}{2}, \end{array}$$

from which we derive, using for the second inequality $\varepsilon \leq r_{k',p'}$ from (2),

$$\begin{array}{rcl} \|s_i e_j - \widehat{c}_{k',p'}\|^2 & > & \widehat{r}^2_{k',p'} + r_{k',p'}\varepsilon + \dfrac{\varepsilon^2}{2} \\ & \text{or} & \\ \|s_i e_j - \widehat{c}_{k',p'}\|^2 & < & \widehat{r}^2_{k',p'} - r_{k',p'}\varepsilon + \dfrac{\varepsilon^2}{2}. \end{array} \tag{9}$$

Finally, from (4) we have

$$r_{k',p'}\varepsilon + \frac{\varepsilon^2}{2} \quad > \quad 2\widehat{r}_{k',p'} + \gamma^2,$$

and, employing this together with (5), we obtain from (9)

$$\|s_i e_j - \widehat{c}_{k',p'}\|^2 > (\widehat{r}_{k',p'} + \gamma)^2 \quad \text{or} \quad \|s_i e_j - \widehat{c}_{k',p'}\|^2 < (\widehat{r}_{k',p'} - \gamma)^2,$$

which holds if and only if

$$\|s_i e_j - \widehat{c}_{k',p'}\| > \widehat{b}_{k',p'} \quad \text{or} \quad \|s_i e_j - \widehat{c}_{k',p'}\| < \widehat{a}_{k',p'}.$$

This shows that $H_{k',p'}$ outputs 0 also in this case. Thus, Claim (iii) is established.

We complete the construction of $\mathcal{N}$ by connecting every hidden node with weight 1 to the output node, which then computes the sum of the hidden node output values.

We finally show that we have indeed obtained a network that induces the dichotomy (S_0, S_1). Assume that $s_i e_j \in S_1$. Claims (i), (ii), and (iii) imply that there is exactly one hidden node $H_{k,p}$, namely one satisfying $k = f(i)$ by the definition of f, that outputs 1 on $s_i e_j$. Hence, the network outputs 1 as well. On the other hand, if $s_i e_j \in S_0$, it follows from Claims (ii) and (iii) that none of the hidden nodes outputs 1. Therefore, the network output is 0. Thus, $\mathcal{N}$ shatters S with threshold 1/2 and the theorem is proven. $\square$

The construction in the previous proof was based on the assumption that the difference between center radius and surround radius, given by the value 2γ, can be made sufficiently small. This may require constraints for the precision of computation that are not available in natural or artificial systems. It is possible, however, to obtain the same result even if there is a lower bound on γ by simply scaling the elements of the shattered set using a sufficiently large factor.

In the following we obtain a superlinear lower bound for the VC dimension of networks with center-surround receptive field neurons. By $\lfloor x \rfloor$ we denote the largest integer less or equal to x.

Corollary 3. *Suppose $\mathcal{N}$ is a network with one hidden layer of k binary CSRF neurons and input dimension $n \geq 2$, where $k \leq 2^n$, and assume that the output node is linear. Then $\mathcal{N}$ has VC dimension at least*

$$\left\lfloor \frac{k}{2} \right\rfloor \cdot \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor \cdot \left(n - \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor + 1\right).$$

This even holds if the weights of the output node are not adjustable.

Proof. We use Theorem 2 with $h = \lfloor k/2 \rfloor$, $q = \lfloor \log(k/2) \rfloor$, and $m = n - \lfloor \log(k/2) \rfloor$. The condition $k \leq 2^n$ guarantees that $m \geq 1$. Then there is a set of cardinality

$$hq(m+1) \;=\; \left\lfloor \frac{k}{2} \right\rfloor \cdot \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor \cdot \left(n - \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor + 1\right).$$

that is shattered by the network specified in Theorem 2. Since the number of hidden nodes is $h + 2^q \leq k$ and the input dimension is $m + q = n$, the network satisfies the required conditions. Furthermore, it was shown in the proof of Theorem 2 that all weights of the output node can be fixed to 1. Hence, they need not be adjustable. $\square$

Corollary 3 immediately implies the following statement, which gives a superlinear lower bound in terms of the number of weights and the number of hidden nodes.

Corollary 4. *Consider a network $\mathcal{N}$ with input dimension $n \geq 2$, one hidden layer of k binary CSRF neurons, where $k \leq 2^{n/2}$, and a linear output node. Let $W = k(n+2) + k + 1$ denote the number of weights. Then $\mathcal{N}$ has VC dimension at least*

$$\frac{W}{5} \log\left(\frac{k}{4}\right).$$

This even holds if the weights of the output node are fixed.

Proof. According to Corollary 3, $\mathcal{N}$ has VC dimension at least $\lfloor k/2 \rfloor \cdot \lfloor \log(k/2) \rfloor \cdot (n - \lfloor \log(k/2) \rfloor + 1)$. The condition $k \leq 2^{n/2}$ implies

$$n - \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor + 1 \;\geq\; \frac{n+4}{2}.$$

We may assume that $k \geq 5$. (The statement is trivial for $k \leq 4$.) It follows, using $\lfloor k/2 \rfloor \geq (k-1)/2$ and $k/10 \geq 1/2$, that

$$\left\lfloor \frac{k}{2} \right\rfloor \;\geq\; \frac{2k}{5}.$$

Finally, we have

$$\left\lfloor \log\left(\frac{k}{2}\right)\right\rfloor \;\geq\; \log\left(\frac{k}{2}\right) - 1 \;=\; \log\left(\frac{k}{4}\right).$$

Hence, $\mathcal{N}$ has VC dimension at least $(n+4)(k/5)\log(k/4)$, which is at least as large as the claimed bound $(W/5)\log(k/4)$. ☐

In the networks considered thus far the input dimension was assumed to be variable. It is an easy consequence of Theorem 2 that even when n is constant, the VC dimension grows still linearly in terms of the network size.

Corollary 5. *Assume that the input dimension is fixed and consider a network $\mathcal{N}$ with one hidden layer of binary CSRF neurons and a linear output node. Then the VC dimension of $\mathcal{N}$ is $\Omega(k)$ and $\Omega(W)$, where k is the number of hidden nodes and W the number of weights. This even holds in the case of fixed output weights.*

Proof. Choose $m, q \geq 1$ such that $m + q \leq n$, and let $h = k - 2^q$. Since n is constant, $hq(m+1)$ is $\Omega(k)$. Thus, according to Theorem 2, there is a set of cardinality $\Omega(k)$ shattered by $\mathcal{N}$. Since the number of weights is $O(k)$, the bound $\Omega(W)$ also follows. ☐

4 Lower Bounds for RBF and DOG Networks

In the following we present the lower bounds for networks with one hidden layer of Gaussian radial basis function neurons and difference-of-Gaussians neurons, respectively. We first consider the latter type.

Theorem 6. *Let $h, q, m \geq 1$ be arbitrary natural numbers. Suppose $\mathcal{N}$ is a network with $m+q$ input nodes, one hidden layer of $h + 2^q$ DOG neurons, and a linear output node. Then there is a set of cardinality $hq(m+1)$ shattered by $\mathcal{N}$.*

Proof. We use ideas and results from the proof of Theorem 2. In particular, we show that the set constructed there can be shattered by a network of new model neurons, the so-called extended Gaussian neurons which we introduce below. Then we demonstrate that a network of these extended Gaussian neurons can be simulated by a network of DOG neurons, which establishes the statement of the theorem.

We define an extended Gaussian neuron with n inputs to compute the function $\tilde{g} : \mathbb{R}^{2n+2} \to \mathbb{R}$ with

$$\tilde{g}(c, \sigma, \alpha, x) \;=\; 1 - \left(\alpha \exp\left(-\frac{\|x - c\|^2}{\sigma^2}\right) - 1\right)^2,$$

where $x_1, \ldots, x_n$ are the input variables, $c_1, \ldots, c_n$, α, and $\sigma > 0$ are real-valued parameters. Thus, the computation of an extended Gaussian neuron is performed by scaling the output of a Gaussian RBF neuron with α, squaring the difference to 1, and comparing this value with 1.

Let $S \subseteq \mathbb{R}^{m+q}$ be the set of cardinality $hq(m+1)$ constructed in the proof of Theorem 2. In particular, S has the form

$$S = \{s_i e_j : i = 1, \ldots, h(m+1); j = 1, \ldots, q\}.$$

We have also defined in that proof binary CSRF neurons $H_{k,p}$ as hidden nodes in terms of parameters $\widehat{c}_{k,p} \in \mathbb{R}^{m+q}$, which became the centers of the neurons, and $\widehat{r}_{k,p} \in \mathbb{R}$, which gave the center radii $\widehat{a}_{k,p} = \widehat{r}_{k,p} - \gamma$ and the surround radii $\widehat{b}_{k,p} = \widehat{r}_{k,p} + \gamma$ using some $\gamma > 0$. The number of hidden nodes was not larger than $h + 2^q$. We replace the CSRF neurons by extended Gaussian neurons $G_{k,p}$ with parameters $c_{k,p}, \sigma_{k,p}, \alpha_{k,p}$ defined as follows. Assume some $\sigma > 0$ that will be specified later. We let

$$\begin{aligned} c_{k,p} &= \widehat{c}_{k,p}, \\ \sigma_{k,p} &= \sigma, \\ \alpha_{k,p} &= \exp\left(\frac{\widehat{r}^{\,2}_{k,p}}{\sigma^2}\right). \end{aligned}$$

These hidden nodes are connected to the output node with all weights being 1. We call this network $\mathcal{N}'$ and claim that it shatters S. Consider some arbitrary dichotomy (S_0, S_1) of S and some $s_i e_j \in S$. Then node $G_{k,p}$ computes

$$\begin{aligned} \tilde{g}(c_{k,p}, \sigma_{k,p}, \alpha_{k,p}, s_i e_j) &= 1 - \left(\alpha_{k,p} \exp\left(-\frac{\|s_i e_j - c_{k,p}\|^2}{\sigma^2_{k,p}}\right) - 1\right)^2 \\ &= 1 - \left(\exp\left(\frac{\widehat{r}^{\,2}_{k,p}}{\sigma^2}\right) \cdot \exp\left(-\frac{\|s_i e_j - \widehat{c}_{k,p}\|^2}{\sigma^2}\right) - 1\right)^2 \\ &= 1 - \left(\exp\left(-\frac{\|s_i e_j - \widehat{c}_{k,p}\|^2 - \widehat{r}^{\,2}_{k,p}}{\sigma^2}\right) - 1\right)^2. \end{aligned} \tag{10}$$

Suppose first that $s_i e_j \in S_1$. It was shown by Claims (i), (ii), and (iii) in the proof of Theorem 2 that there is exactly one hidden node $H_{k,p}$ that outputs 1 on $s_i e_j$. In particular, the proof of Claim (i) established that this node satisfies

$$\|s_i e_j - \widehat{c}_{k,p}\|^2 = \widehat{r}^{\,2}_{k,p}.$$

Hence, according to (10) node $G_{k,p}$ outputs 1. We note that this holds for all values of σ. Further, the proofs of Claims (ii) and (iii) yielded that those nodes $H_{k,p}$ that output 0 on $s_i e_j$ satisfy

$$\|s_i e_j - \widehat{c}_{k,p}\|^2 > (\widehat{r}_{k,p} + \gamma)^2 \quad \text{or} \quad \|s_i e_j - \widehat{c}_{k,p}\|^2 < (\widehat{r}_{k,p} - \gamma)^2.$$

This implies for the computation of $G_{k,p}$ that in (10) we can make the expression

$$\exp\left(-\frac{\|s_i e_j - \widehat{c}_{k,p}\|^2 - \widehat{r}_{k,p}^2}{\sigma^2}\right)$$

as close to 0 as necessary by choosing σ sufficiently small. Since this does not affect the node that outputs 1, network $\mathcal{N}'$ computes a value close to 1 on $s_i e_j$.

On the other hand, for the case $s_i e_j \in S_0$ it was shown in Theorem 2 that all nodes $H_{k,p}$ output 0. Thus, if σ is sufficiently small, each node $G_{k,p}$, and hence $\mathcal{N}'$, outputs a value close to 0. Hence, S is shattered by thresholding the output of $\mathcal{N}'$ at 1/2.

Finally, we show that S can be shattered by a network $\mathcal{N}$ of the same size with DOG neurons as hidden nodes. The computation of an extended Gaussian neuron can be rewritten as

$$\begin{aligned}
\tilde{g}(c,\sigma,\alpha,x) &= 1 - \left(\alpha \exp\left(-\frac{\|x-c\|^2}{\sigma^2}\right) - 1\right)^2 \\
&= 1 - \left(\alpha^2 \exp\left(-\frac{2\|x-c\|^2}{\sigma^2}\right) - 2\alpha \exp\left(-\frac{\|x-c\|^2}{\sigma^2}\right) + 1\right) \\
&= 2\alpha \exp\left(-\frac{\|x-c\|^2}{\sigma^2}\right) - \alpha^2 \exp\left(-\frac{2\|x-c\|^2}{\sigma^2}\right) \\
&= g_{\mathrm{DOG}}(c,\sigma,\sigma/\sqrt{2},2\alpha,\alpha^2,x).
\end{aligned}$$

Hence, the extended Gaussian neuron is equivalent to a weighted difference of two Gaussian neurons with center c, widths $\sigma, \sigma/\sqrt{2}$ and weights $2\alpha, \alpha^2$, respectively. Thus, the extended Gaussian neurons can be replaced by the same number of DOG neurons. □

We note that the network of extended Gaussian neurons constructed in the previous proof has all output weights fixed, whereas the output weights of the DOG neurons, that is, the parameters α and β in the notation of Section 2, are calculated from the parameters of the extended Gaussian neurons and, therefore, depend on the particular dichotomy to be implemented. (It is trivial for a DOG network to have an output node with fixed weights since the DOG neurons have built-in output weights.)

We are now able to deduce a superlinear lower bound on the VC dimension of DOG networks.

Corollary 7. *Suppose $\mathcal{N}$ is a network with one hidden layer of DOG neurons and a linear output node. Let $\mathcal{N}$ have k hidden nodes and input dimension $n \geq 2$, where $k \leq 2^n$. Then $\mathcal{N}$ has VC dimension at least*

$$\left\lfloor \frac{k}{2} \right\rfloor \cdot \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor \cdot \left(n - \left\lfloor \log\left(\frac{k}{2}\right) \right\rfloor + 1\right).$$

Let W denote the number of weights and assume that $k \leq 2^{n/2}$. Then the VC dimension of $\mathcal{N}$ is at least

$$\frac{W}{5} \log \left(\frac{k}{4} \right) .$$

For fixed input dimension the VC dimension of $\mathcal{N}$ is bounded by $\Omega(k)$ and $\Omega(W)$.

Proof. The results are implied by Theorem 6 in the same way as Corollaries 3, 4, and 5 follow from Theorem 2. □

Finally, we have the lower bound for Gaussian RBF networks.

Theorem 8. *Suppose $\mathcal{N}$ is a network with one hidden layer of Gaussian RBF neurons and a linear output node. Let k be the number of hidden nodes and n the input dimension, where $n \geq 2$ and $k \leq 2^{n+1}$. Then $\mathcal{N}$ has VC dimension at least*

$$\left\lfloor \frac{k}{4} \right\rfloor \cdot \left\lfloor \log \left(\frac{k}{4} \right) \right\rfloor \cdot \left(n - \left\lfloor \log \left(\frac{k}{4} \right) \right\rfloor + 1 \right) .$$

Let W denote the number of weights and assume that $k \leq 2^{(n+2)/2}$. Then the VC dimension of $\mathcal{N}$ is at least

$$\frac{W}{12} \log \left(\frac{k}{8} \right) .$$

For fixed input dimension $n \geq 2$ the VC dimension of $\mathcal{N}$ satisfies the bounds $\Omega(k)$ and $\Omega(W)$.

Proof. Clearly, a DOG neuron can be simulated by two weighted Gaussian RBF Neurons. Thus, by virtue of Theorem 6 there is a network $\mathcal{N}$ with $m + q$ input nodes and one hidden layer of $2(h + 2^q)$ Gaussian RBF neurons that shatters some set of cardinality $hq(m + 1)$. Choosing $h = \lfloor k/4 \rfloor, q = \lfloor \log(k/4) \rfloor$, and $m = n - \lfloor \log(k/4) \rfloor$ we obtain similarly to Corollary 3 the claimed lower bound in terms of n and k. Furthermore, the stated bound in terms of W and k follows by analogy to Corollary 4. The bound for fixed input dimension is obvious, as in the proof of Corollary 5. □

Some radial basis function networks studied theoretically or used in practice have no adjustable width parameters (for instance in [5, 20]). Therefore, a natural question is whether the previous result also holds for networks with fixed width parameters. The values of the width parameters for Theorem 8 arise from the widths of DOG neurons specified in Theorem 6. The two width parameters of each DOG neuron have the form σ and $\sigma/\sqrt{2}$ where σ is common to all DOG neurons and is only required to be sufficiently small. Hence, we can choose a single σ that is sufficiently small for all dichotomies to be induced. Thus, for the RBF network we not only have that the width parameters can be fixed, but even that there need to be only two different width values—solely depending on the architecture and not on the particular dichotomy.

Corollary 9. *Let $\mathcal{N}$ be a Gaussian RBF network with n input nodes and k hidden nodes satisfying the conditions of Theorem 8. Then there exists a real number $\sigma_{n,k} > 0$ such that the VC dimension bounds stated in Theorem 8 hold for $\mathcal{N}$ with each RBF neuron having fixed width $\sigma_{k,n}$ or $\sigma_{k,n}/\sqrt{2}$.*

With regard to Theorem 8 we further remark that k has been previously established as lower bound for RBF networks by Anthony and Holden [2]. Further, also Theorem 19 of Lee et al. [11] in connection with the result of Erlich et al. [7] implies the lower bound $\Omega(nk)$, and hence $\Omega(k)$ for fixed input dimension. By means of Theorem 8 we are now able to present a lower bound that is even superlinear in k.

Corollary 10. *Let $n \geq 2$ and $\mathcal{N}$ be the network with $k = 2^{n+1}$ hidden Gaussian RBF neurons. Then $\mathcal{N}$ has VC dimension at least*

$$\frac{k}{3}\log\left(\frac{k}{8}\right).$$

Proof. Since $k = 2^{n+1}$, we may substitute $n = \log k - 1$ in the first bound of Theorem 8. Hence, the VC dimension of $\mathcal{N}$ is at least

$$\left\lfloor\frac{k}{4}\right\rfloor \cdot \left\lfloor\log\left(\frac{k}{4}\right)\right\rfloor \cdot \left(\log k - \left\lfloor\log\left(\frac{k}{4}\right)\right\rfloor\right) \;\; \geq \;\; 2\left\lfloor\frac{k}{4}\right\rfloor \cdot \left\lfloor\log\left(\frac{k}{4}\right)\right\rfloor .$$

Using $\lfloor k/4 \rfloor \geq k/6$ and $\lfloor \log(k/4) \rfloor \geq \log(k/8)$ yields the claimed bound. □

5 Concluding Remarks

We have shown that the VC dimension of every reasonably sized one-hidden-layer network of RBF, DOG, and binary CSRF neurons is superlinear. It is not difficult to deduce that the bound for binary CSRF networks is asymptotically tight. For RBF and DOG networks, however, the currently available methods give only rise to the upper bound $O(W^2k^2)$. To narrow the gap between upper and lower bounds for these networks is an interesting open problem.

It is also easy to obtain a linear upper bound for the single neuron in the RBF and binary CSRF case, whereas for the DOG neuron the upper bound is quadratic. We conjecture that also the DOG neuron has a linear VC dimension, but the methods currently available do not seem to permit an answer.

The bounds we have derived involve constant factors that are the largest known for any standard neural network with one hidden layer. This fact could be evidence of the higher cooperative computational capabilities of local receptive field neurons in comparison to other neuron types. This statement, however, must be taken with care since the constants involved in the bounds are not yet known to be tight.

RBF neural networks compute a particular type of kernel-based functions. The method we have developed for obtaining the results presented here is of quite general nature. We expect it therefore to be applicable for other kernel-based function classes as well.

Acknowledgment

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

References

[1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, Cambridge, 1999.

[2] M. Anthony and S. B. Holden. Quantifying generalization in linearly weighted neural networks. *Complex Systems*, 8:91–114, 1994.

[3] P. L. Bartlett, V. Maiorov, and R. Meir. Almost linear VC dimension bounds for piecewise polynomial networks. *Neural Computation*, 10:2159–2173, 1998.

[4] C. M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

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

[6] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 14:326–334, 1965.

[7] Y. Erlich, D. Chazan, S. Petrack, and A. Levy. Lower bound on VC-dimension by local shattering. *Neural Computation*, 9:771–776, 1997.

[8] V. D. Glezer. *Vision and Mind: Modeling Mental Functions.* Lawrence Erlbaum, Mahwah, New Jersey, 1995.

[9] E. J. Hartman, J. D. Keeler, and J. M. Kowalski. Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation*, 2:210–215, 1990.

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

[11] W. S. Lee, P. L. Bartlett, and R. C. Williamson. Lower bounds on the VC dimension of smoothly parameterized function classes. *Neural Computation*, 7:1040–1053, 1995.

[12] W. S. Lee, P. L. Bartlett, and R. C. Williamson. Correction to "Lower bounds on VC-dimension of smoothly parameterized function classes". *Neural Computation*, 9:765–769, 1997.

[13] W. Maass. Neural nets with super-linear VC-dimension. *Neural Computation*, 6:877–884, 1994.

[14] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman, New York, 1982.

[15] H. N. Mhaskar. Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation*, 8:164–177, 1996.

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

[17] J. G. Nicholls, A. R. Martin, and B. G. Wallace. *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System.* Sinauer Associates, Sunderland, Mass., third edition, 1992.

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

[19] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78:1481–1497, 1990.

[20] M. J. D. Powell. The theory of radial basis function approximation in 1990. In W. Light, editor, *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms, and Radial Basis Functions*, chapter 3, pages 105–210. Clarendon Press, Oxford, 1992.

[21] B. D. Ripley. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, 1996.

[22] A. Sakurai. Tighter bounds of the VC-dimension of three layer networks. In *Proceedings of the World Congress on Neural Networks*, volume 3, pages 540–543. Erlbaum, Hillsdale, New Jersey, 1993.

[23] M. Tessier-Lavigne. Phototransduction and information processing in the retina. In E. R. Kandel, J. H. Schwartz, and T. M. Jessell, editors, *Principles of Neural Science*, chapter 28, pages 400–418. Prentice Hall, Englewood Cliffs, New Jersey, third edition, 1991.