Annals of Mathematics and Artificial Intelligence (2024) 92:691–701
https://doi.org/10.1007/s10472-023-09901-x

RESEARCH

On the universal approximation property of radial basis function neural networks

Aysu Ismayilova[1] · Muhammad Ismayilov[2]

Accepted: 3 October 2023 / Published online: 16 October 2023

Abstract

In this paper we consider a new class of RBF (Radial Basis Function) neural networks, in which smoothing factors are replaced with shifts. We prove under certain conditions on the activation function that these networks are capable of approximating any continuous multivariate function on any compact subset of the d-dimensional Euclidean space. For RBF networks with finitely many fixed centroids we describe conditions guaranteeing approximation with arbitrary precision.

Keywords RBF neural network · Activation function · Mean-periodic function · Centroid · Shift

Mathematics Subject Classification (2010) 41A30 · 41A63 · 68T05 · 92B20

1 Introduction

RBF (Radial Basis Function) neural networks are being used for function approximation, time series forecasting, classification, pattern recognition and system control problems. Besides their strong approximation capability, these networks benefit from other powerful characteristics such as the ability to represent complex nonlinear mappings and provide a fast and robust learning mechanism without significant computational cost. The literature abounds with different aspects and various applications of RBF neural networks. For instance, Agarwal et al. [1] employed RBF neural networks in face recognition, adapting hidden neuron centers using a Firefly Algorithm (FA). This approach resulted in improved recognition accuracy. The higher performance of the proposed technique in face recognition was shown by experimental validation on well recognized face datasets, when compared to existing methods. Khan et al. [9] employed the MATLAB neural network toolbox alongside RBF neural networks

✉ Muhammad Ismayilov
mv.ismayilov@gmail.com

Aysu Ismayilova
a.ismayilova-1@sms.ed.ac.uk

1 School of Mathematics, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, Scotland

2 French-Azerbaijani University, 183 Nizami str., Baku, Azerbaijan

to predict corrosion-induced failures in subsurface oil and gas pipelines. They utilized fault tree analysis to identify corrosion-related hazards. Through training, the RBF neural network demonstrated accuracy in forecasting the likelihood of underground pipeline corrosion failures. Wang et al. [23] used RBF neural networks to improve the tracking stability control of Unmanned Surface Vehicles (USVs). Wang et al. [22] used RBF neural networks for image reconstruction in Electric Impedance Tomography (EIT) and achieved high accuracy. Fath et al. [4] employed RBF neural networks in oil and gas industry. They developed reliable models based on multilayer perceptron (MLP) and RBF neural networks for estimating the solution gas-oil ratios. These models were tested using a total of 710 experiments across various conditions. Both models outperformed existing empirical correlations, with the RBF neural network proving more efficient and accurate, offering a valuable tool for accurate calculations of gas-oil ratios. Karamichailidou et al. [8] employed RBF neural network as a pivotal tool in modeling wind turbine performance. This innovative approach takes into account various environmental factors, such as wind direction, ambient temperature, and blade pitch angle, to develop precise power curve models. Li et al. [10] achieved significant advancements in credit rating modeling by augmenting RBF neural networks with an optimized segmentation algorithm. Their research led to improved model robustness, accuracy, and adaptability, with a particular emphasis on handling non-numeric data effectively. These enhancements have important implications for the field of credit risk management in the banking and finance sector.

The fundamental principles and advantages of RBF neural networks were first displayed in the papers of Broomhead and Lowe [3], Moody and Darken [14], Lipmann [12] and Bishop [2]. A variant of RBF network with an input layer, a hidden layer and an output layer is constructed by the following scheme. Each unit in the hidden layer of this RBF network has its own centroid and for an input vector $\mathbf{x} = (x_1, ..., x_d)$ it computes the distance between $\mathbf{x}$ and its centroid $\mathbf{c} \in \mathbb{R}^d$. Its output (the output of a given hidden unit) is some nonlinear function of that distance. Hence each hidden unit computes a radial function, that is, a function which is constant on the spheres $\|\mathbf{x} - \mathbf{c}\| = \alpha,\ \alpha \in \mathbb{R}$. Each output unit gives a weighted summation of the outputs of hidden units. For the clarity of exposition, we will consider in the sequel only a one dimensional output space instead of outputs represented by multiple units. The generalization of our results to the n-dimensional output space is straightforward.

Assuming that there are d input units and one output unit, the final response function has the following form:

$$G(\mathbf{x}) = \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma_i}\right). \tag{1.1}$$

Here $m \in \mathbb{N}$ is the number of units in the hidden layer, $(w_1, ..., w_m) \in \mathbb{R}^m$ is the vector of weights, $\mathbf{x} \in \mathbb{R}^d$ is an input vector, $\mathbf{c}_i \in \mathbb{R}^d$ and $\sigma_i \in \mathbb{R}_+$ are the centroids and smoothing factors (or widths) of the i-th node, $1 \leq i \leq m$, respectively, $\|\mathbf{x} - \mathbf{c}_i\|$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{c}_i$, and $g : [0, +\infty) \to \mathbb{R}$ is the so-called activation function.

Various activation functions in RBF neural networks can be implemented and the smoothing factors may be the same or may vary across units.

The RBF neural networks have the universal approximation property. Theoretically, such networks can approximate any continuous multivariate function within any degree of accuracy, if the activation function is suitably chosen. The most well-known result is due to Park and Sandberg [15]. In 1993, they showed, along with other results, that for a continuous and integrable $g(\|\mathbf{x}\|)$ (considered as a function of d variables) the set of functions (1.1) is dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compact subsets of $\mathbb{R}^d$. That is, for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$, for any compact subset $K \subset \mathbb{R}^d$ and for any $\varepsilon > 0$, there exists a function G of form (1.1) such that

$$\|f - G\|_K \overset{def}{=} \max_{\mathbf{x} \in K} |f(\mathbf{x}) - G(\mathbf{x})| < \varepsilon.$$

The requirement of the integrability of $g(\|\mathbf{x}\|)$ is relaxed in Liao, Fang and Nuttle [11]. They showed that for an activation function, which is continuous almost everywhere, locally essentially bounded and nonpolynomial, the RBF networks (1.1) can approximate any continuous function with arbitrary accuracy. There are also other results on the universality of RBF neural networks (see, e.g., [6, 11, 16, 24]).

In this paper we bring into consideration a new class of RBF neural networks. In this class the smoothing factors σ_i are replaced with shifts $\nu_i \in \mathbb{R}$. That is, this class consists of functions $H : \mathbb{R}^d \to \mathbb{R}$ of the form

$$H(\mathbf{x}) = \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right). \tag{1.2}$$

We are interested in the universal approximation property of such RBF neural networks. For which activations g, functions of form (1.2) are dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compact subsets of $\mathbb{R}^d$. We will give various conditions on the activation g which guarantee the density of the functions (1.2) in $C(X)$ for any compact set $X \subset \mathbb{R}^d$.

The utilization of RBF neural networks with shifts represents a novel approach that has yet to find practical applications. However, because of its flexibility and possible versatility, this novel technique is promising in a variety of fields. Note that by using shifts, one effectively allows the activation function g to be translated to different positions, which may have extra advantages. The use of shifts can make the RBF network more flexible in capturing complex patterns in the data. It allows the model to adapt better to irregularly shaped data distributions.

2 Universal approximation theorems

The following theorem is based on the results of Park and Sandberg [15], and Schwartz [18].

Theorem 2.1 *Assume* $d \geq 1$, $1 \leq p < \infty$. *Assume an activation function* $g \in C(\mathbb{R}) \cap L^p(\mathbb{R})$ *and* $t^{d-1} g(t)$ *is integrable on* $[0, +\infty)$. *Then for any continuous function* $f : \mathbb{R}^d \to \mathbb{R}$, *for any* $\varepsilon > 0$ *and for any compact subset* $X \subset \mathbb{R}^d$, *there exist* $m \in \mathbb{N}$, $w_i, \nu_i \in \mathbb{R}$, $\mathbf{c}_i \in \mathbb{R}^d$ *such that for all* $\mathbf{x} \in X$

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right) \right| < \varepsilon. \tag{2.1}$$

Proof The result of Park and Sandberg (see [15, Theorem 5]) says that if $K : \mathbb{R}^d \to \mathbb{R}$ is continuous, integrable and radially symmetric with respect to the Euclidean norm, then the functions of the form

$$q(\mathbf{x}) = \sum_{i=1}^{k} w_i K\left(\frac{\mathbf{x} - \mathbf{c}_i}{\sigma_i}\right). \tag{2.2}$$

are dense in $C(X)$ for any compact set $X \subset \mathbb{R}^d$.

Consider the function $K(\mathbf{x}) = g(\|\mathbf{x}\|)$. Clearly, $K(\mathbf{x})$ is radially symmetric and since $t^{d-1}g(t)$ is integrable on $[0, +\infty)$, $K(\mathbf{x})$ is integrable on $\mathbb{R}^d$. Thus this function satisfies the hypothesis of Park and Sandberg's theorem. Assume any function $f \in C(\mathbb{R}^d)$, any number $\varepsilon > 0$ and any compact subset $X \subset \mathbb{R}^d$ are given. By the above result of Park and Sandberg, there exist $k \in \mathbb{N}$, $w_i \in \mathbb{R}$, $\sigma_i > 0$, $\mathbf{c}_i \in \mathbb{R}^d$ such that

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma_i}\right)\right| < \frac{\varepsilon}{2} \tag{2.3}$$

for all $\mathbf{x} \in X$. Note that we can write inequality (2.3) in the form

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} g_i\left(\|\mathbf{x} - \mathbf{c}_i\|\right)\right| < \frac{\varepsilon}{2}, \tag{2.4}$$

where $g_i(t) := w_i g(t/\sigma_i)$, $t \in \mathbb{R}$. Since X is compact and the distance function $\|\cdot\|$ is continuous, the sets $\{\|\mathbf{x} - \mathbf{c}_i\| : \mathbf{x} \in X\}$ are compact subsets of $\mathbb{R}$; hence $\{\|\mathbf{x} - \mathbf{c}_i\| : \mathbf{x} \in X\} \subset [a_i, b_i]$ for some finite a_i and b_i, $i = 1, ..., k$.

In 1947, Schwartz [18] proved that continuous and p-th degree ($1 \le p < \infty$) Lebesgue integrable univariate functions are not mean-periodic (see also [17, Proposition 3.12]). Note that a function $u \in C(\mathbb{R}^d)$ is called mean periodic if the set $span\ \{u(\mathbf{x} - \mathbf{a}) :\ \mathbf{a} \in \mathbb{R}^d\}$ is not dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compacta (see [18]). Since $g \in C(\mathbb{R}) \cap L^p(\mathbb{R})$, by this result of Schwartz, the set

$$span\ \{g(t - \lambda) :\ \lambda \in \mathbb{R}\}$$

is dense in $C(\mathbb{R})$ in the topology of uniform convergence. This density result means that for the given ε there exist numbers $\rho_{ij}, \lambda_{ij} \in \mathbb{R}$, $i = 1, 2, ..., k$, $j = 1, ..., s_i$ such that

$$\left| g_i(t) - \sum_{j=1}^{s_i} \rho_{ij} g(t - \lambda_{ij})\right| < \frac{\varepsilon}{2k} \tag{2.5}$$

for all $t \in [a_i, b_i]_i,\ i = 1, 2, ..., k$. From (2.4) and (2.5) it follows that

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k}\sum_{j=1}^{s_i} \rho_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \lambda_{ij})\right| < \varepsilon, \tag{2.6}$$

for all $\mathbf{x} \in X$. After writing the double sum in (2.6) as a single sum we obtain the validity of (2.1).

Corollary 2.1 *Assume g is a continuous, monotone and bounded function on $\mathbb{R}$ and $t^{d-1}g(t)$ is integrable on $[0, +\infty)$. Then for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$, for any $\varepsilon > 0$ and for any compact subset $X \subset \mathbb{R}^d$, there exist $m \in \mathbb{N}$, $w_i, \nu_i \in \mathbb{R}$, $\mathbf{c}_i \in \mathbb{R}^d$ such that*

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right)\right| < \varepsilon$$

for all $\mathbf{x} \in X$.

Proof In [5] Funahashi proved that if g is a continuous, monotone and bounded function on $\mathbb{R}$, then the function $h(t) = g(t + \alpha) - g(t - \alpha)$ belongs to $L_1(\mathbb{R})$ for any real α. Thus the function $h(t)$ is not mean periodic being a continuous and L_1 function. In addition, $t^{d-1}h(t)$

is integrable on $[0, +\infty)$. We can apply the above theorem to h and then changing $h(\cdot)$ to $g(\cdot+\alpha) - g(\cdot-\alpha)$ obtain the desired result.

In the following theorem, integrability condition is not required.

Theorem 2.2 *Assume g is a nonconstant continuous bounded function on $\mathbb{R}$ which has a limit at infinity or minus infinity. Then for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$, for any $\varepsilon > 0$ and for any compact subset $X \subset \mathbb{R}^d$, there exist $m \in \mathbb{N}$, $w_i, \nu_i \in \mathbb{R}$, $\mathbf{c}_i \in \mathbb{R}^d$ such that for all $\mathbf{x} \in X$*

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right) \right| < \varepsilon.$$

Proof The conditions on g implies that the function $K_0(\mathbf{x}) = g(\|\mathbf{x}\|)$ is a nonconstant, continuous, bounded multivariate function. Hence $K_0(\mathbf{x})$ is not a polynomial. Here we use Liao, Fang and Nuttle's result from [11], where they proved that for a function $K(\mathbf{x})$, which is continuous almost everywhere, locally essentially bounded and nonpolynomial, the RBF networks $\sum_{i=1}^{m} w_i K\left(\frac{\mathbf{x}-\mathbf{c}_i}{\sigma_i}\right)$ are dense in $C(X)$ for any compact set $X \subset \mathbb{R}^d$. Note that our function $K_0(\mathbf{x})$ satisfy conditions of this result. In addition, it follows from one theorem of Schwartz (see [18, p.907] and [17, Proposition 3.12]) that g is not mean-periodic. Therefore, the set $span\{g(t-\lambda) : \lambda \in \mathbb{R}\}$ is dense in $C[a, b]$ for any closed interval $[a, b]$. Now we have all the necessary facts to repeat the above ideas from the proof of Theorem 2.1 and obtain the desired final result.

Remark Corollary 2.1 can be obtained from Theorem 2.2 directly, without using Funahashi's above result on the Lebesgue integrability of the function $h(t) = g(t+\alpha) - g(t-\alpha)$.

3 RBF networks with finitely many centroids

In this section we study approximation properties of RBF neural networks that have a finite number of fixed centroids. We describe compact sets $X \subset \mathbb{R}^d$, for which such networks are dense in $C(X)$.

Assume we are given k fixed centroids $\mathbf{c}_1, ..., \mathbf{c}_k$. Put $S = \{\mathbf{c}_1, ..., \mathbf{c}_k\}$. Consider the set of RBF networks with only these centroids and arbitrary shifts

$$\mathcal{G}(g, S) = \left\{ \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}\| - \nu_i\right) : \mathbf{c} \in S,\ w_i, \nu_i \in \mathbb{R},\ m \in \mathbb{N} \right\}.$$

In the above set we fix only the set $S = \{\mathbf{c}_1, ..., \mathbf{c}_k\}$ and the activation function $g \in C(\mathbb{R})$.

The difference between the set of RBF networks $H(x)$ (1.2) implemented in the previous section and $\mathcal{G}(g, S)$ is as follows. In $H(x)$, all the parameters, including weights, centers, and shifts, can vary independently for each term, allowing for flexibility and adaptation to different situations. In contrast, functions within $\mathcal{G}(g, S)$ share a fixed set of centers S. While weights and shifts can differ among functions, the centers remain constant across all functions in the set, providing a common reference point. Fixing centers makes the network training simpler and computationally less demanding. This is because one does not need to estimate center locations during training, which can be particularly advantageous when dealing with limited data and real-time applications.

The following questions arise naturally:

1) Are the RBF networks from $\mathcal{G}(g, S)$ dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compacta. That is, if for any compact set X, $\overline{\mathcal{G}(g, S)} = C(X)$?

2) If the answer to the above question is negative, for which compact sets $X \subset \mathbb{R}^d$, $\overline{\mathcal{G}(g, S)} = C(X)$?

Note that the answer to the 1st question is indeed negative. To see this, introduce the following set of functions:

$$\mathcal{R}(S) = \left\{ \sum_{i=1}^{k} g_i \left(\|\mathbf{x} - \mathbf{c}_i\| \right) : g_i \in C(\mathbb{R}) \right\}.$$

In this set $S = \{\mathbf{c}_1, ..., \mathbf{c}_k\}$ is fixed and we vary continuous functions g_i. Note that this is a linear space. Since every summand $w_i g \left(\|\mathbf{x} - \mathbf{c}\| - v_i \right)$ in $\mathcal{G}(g, S)$ is a function of the form $g_i \left(\|\mathbf{x} - \mathbf{c}_i\| \right)$ for some $\mathbf{c}_i$, we deduce that $\mathcal{G}(g, S) \subset \mathcal{R}(S)$. Thus, the set of RBF networks with fixed centroids is smaller than the set of linear combinations of radial functions with that centroids. Therefore, if $\mathcal{G}(g, S)$ was dense in $C(X)$, the set $\mathcal{R}(S)$ would be dense as well. But unfortunately, $\mathcal{R}(S)$ is not dense in $C(X)$ for exceedingly many compact sets X. The reason for the lack of density here is related to the following theorem, which is due to Vitushkin and Henkin [21]: For any k fixed continuously differentiable functions h_i, $i = 1, ..., k$, defined on a cube $[a, b]^d$ the set of functions

$$\left\{ \sum_{i=1}^{k} g_i \left(h_i(x_1, ..., x_d) \right) : g_i \in C(\mathbb{R}) \right\}$$

is nowhere dense in the space of all continuous functions on $[a, b]^d$ with the topology of uniform convergence. Therefore, $\mathcal{G}(g, S)$ cannot be dense in $C(X)$ if all compact sets X are involved. For example, since any set with interior contains a sufficiently small cube $[a, b]^d$, it follows from the result of Vitushkin and Henkin that $\mathcal{G}(g, S)$ is not dense in $C(X)$ for any compact set X with interior points. But there still may be compact sets X for which $\overline{\mathcal{G}(g, S)} = C(X)$ (Take, for example, a single point set $X = \{\mathbf{x}\}$) How can we characterize such sets? To answer this question we introduce the following objects called *cycles*:

Definition 3.1 A set of points $l = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^d$ is called a cycle with respect to the centroids $\mathbf{c}_1, ..., \mathbf{c}_k$ if there exists a vector $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$ such that

$$\sum_{j=1}^{n} \lambda_j \delta_{\|\mathbf{x}_j - \mathbf{c}_i\|}(t) = 0, \quad for\ all\ i = 1, \ldots, k. \tag{3.1}$$

In the above definition $\delta_{\|\mathbf{x}_j - \mathbf{c}_i\|}(t)$ is the characteristic function of the single point set $\{\left\|\mathbf{x}_j - \mathbf{c}_i\right\|\}$. That is,

$$\delta_{\|\mathbf{x}_j - \mathbf{c}_i\|}(t) = \begin{cases} 1, & \text{if } t = \left\|\mathbf{x}_j - \mathbf{c}_i\right\| \\ 0, & \text{if } t \neq \left\|\mathbf{x}_j - \mathbf{c}_i\right\| \end{cases}.$$

Let us look at Eq. (3.1) more closely. We will see that in fact it stands for a system of simple linear equations. To understand this, fix the subscript i. Let the set $\{\left\|\mathbf{x}_j - \mathbf{c}_i\right\|), j = 1, ..., n\}$ have s_i different values, which we denote by $\gamma_1^i, \gamma_2^i, ..., \gamma_{s_i}^i$. Take the first number γ_1^i. Putting $t = \gamma_1^i$, we obtain from (3.1) that

$$\sum_j \lambda_j = 0,$$

where the sum is taken over all j such that $\|\mathbf{x}_j - \mathbf{c}_i\| = \gamma_1^i$. This is the first linear equation in $\lambda_1, ..., \lambda_n$. This equation corresponds to γ_1^i. Take now γ_2^i. By the same way, putting $t = \gamma_2^i$ in (3.1), we can form the second equation. Continuing until $\gamma_{s_i}^i$, we obtain s_i linear homogeneous equations in $\lambda_1, ..., \lambda_n$. The coefficients of these equations are the integers 0 and 1. By varying i, we finally obtain $s = \sum_{i=1}^k s_i$ such equations. Thus we see that (3.1), in its expanded form, stands for the system of these linear equations. Thus the set $l = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is a cycle if this system has a solution with nonzero integer components. In fact, it is not difficult to understand that if the system (3.1) has a solution with nonzero real components, then it has also a solution with nonzero integer components. This means that in the above definition we can replace $\mathbb{Z}^n \setminus \{\mathbf{0}\}$ with $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

We provide two simple examples of cycles here. The reader can give many other examples easily. Assume two centroids $\mathbf{c}_1 = (0, 0)$ and $\mathbf{c}_2 = (4, 0)$ are given in the xy plane. Then any two points A and B on the straight line $x = 2$, which are also symmetric to the line $y = 0$, form a cycle. Indeed, the distances from A and B to $\mathbf{c}_1$ are equal and Eq. (3.1) in case of $i = 1$ will be $\lambda_1 + \lambda_2 = 0$. Since the distances from A and B to $\mathbf{c}_2$ are also equal, Eq. (3.1) yields the same equation for $i = 2$. Thus $\{A, B\}$ is a 2-point cycle and the vector (λ_1, λ_2) can be taken as $(-1, 1)$. It is also easy to construct a 4-point cycle with respect to these centroids. Consider four circles $\|\mathbf{x} - \mathbf{c}_1\| = 2$, $\|\mathbf{x} - \mathbf{c}_1\| = 3$, $\|\mathbf{x} - \mathbf{c}_2\| = 4$, $\|\mathbf{x} - \mathbf{c}_2\| = 3$ in the given order. These circles meet at 4 points A, B, C, D in the 1-st quarter of the xy plane. Each circle has only two of these points and it is not difficult to verify that Eq. (3.1) turns into the system

$$\begin{cases} \lambda_1 + \lambda_2 = 0 \\ \lambda_3 + \lambda_4 = 0 \\ \lambda_2 + \lambda_3 = 0 \\ \lambda_1 + \lambda_4 = 0 \end{cases}$$

which has a solution $(-1, 1 - 1, 1)$. Hence, $\{A, B, C, D\}$ is a cycle with respect to the centroids $\mathbf{c}_1$ and $\mathbf{c}_2$.

The second example above inspires consideration of general cycles with respect to any given two centroids $\mathbf{c}_1$ and $\mathbf{c}_2$. In this special case, we will use the term *closed path* instead of *cycle*.

Definition 3.2 Assume $l = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ with $\mathbf{x}_i \neq \mathbf{x}_{i+1}$, is an ordered set with the property that $\|\mathbf{x}_1 - \mathbf{c}_1\| = \|\mathbf{x}_2 - \mathbf{c}_1\|$, $\|\mathbf{x}_2 - \mathbf{c}_2\| = \|\mathbf{x}_3 - \mathbf{c}_2\|$, $\|\mathbf{x}_3 - \mathbf{c}_1\| = \|\mathbf{x}_4 - \mathbf{c}_1\|$, ... or $\|\mathbf{x}_1 - \mathbf{c}_2\| = \|\mathbf{x}_2 - \mathbf{c}_2\|$, $\|\mathbf{x}_2 - \mathbf{c}_1\| = \|\mathbf{x}_3 - \mathbf{c}_1\|$, $\|\mathbf{x}_3 - \mathbf{c}_2\| = \|\mathbf{x}_4 - \mathbf{c}_2\|$, ... Then l is called a path with respect to the centroids $\mathbf{c}_1$ and $\mathbf{c}_2$. A path having an even number of points $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n})$ is said to be closed if $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n}, \mathbf{x}_1)$ is also a path.

Note that a closed path is a cycle. Indeed if $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n})$ is a closed path, then it is not difficult to see that for a vector $\lambda = (\lambda_1, \dots, \lambda_{2n})$ with the components $\lambda_i = (-1)^i$, we have

$$\sum_{j=1}^{2n} \lambda_j \delta_{\|\mathbf{x}_j - \mathbf{c}_1\|} = 0,$$

$$\sum_{j=1}^{2n} \lambda_j \delta_{\|\mathbf{x}_j - \mathbf{c}_2\|} = 0.$$

Thus, by Definition 3.1, the set $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n}\}$ forms a cycle with respect to the centroids $\mathbf{c}_1$ and $\mathbf{c}_2$.

Cycles and paths may be defined not only for distance functions $d(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}\|$ but also for other useful functions too. There is a rich history of these objects defined for inner products $\mathbf{a} \cdot \mathbf{x}$, which were proved to be very efficient in the theory of ridge functions (that is, functions of the form $g(\mathbf{a} \cdot \mathbf{x})$). See, for example, the monograph by Ismailov [7].

Let $\mathbf{c}_1$, ..., $\mathbf{c}_k$ be fixed centroids and X be a compact subset of $\mathbb{R}^d$. For each $i = 1, \ldots, k$, consider the following set functions

$$\tau_i : 2^X \to X,\ \tau_i(Z) = \{\mathbf{x} \in Z :\ |d_i^{-1}(d_i(\mathbf{x})) \bigcap Z| \geq 2\},$$

where $d_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}_i\|$ and the symbol $|\ \ |$ denotes the cardinality of a considered set. Define $\tau(Z)$ to be $\bigcap_{i=1}^k \tau_i(Z)$ and define $\tau^2(Z) = \tau(\tau(Z))$, $\tau^3(Z) = \tau(\tau^2(Z))$ and so on inductively. Clearly, $\tau(Z) \supseteq \tau^2(Z) \supseteq \tau^3(Z) \supseteq$ ...It is possible that for some n, $\tau^n(Z) = \emptyset$. In this case, one can see that Z does not contain a cycle. In general, if some set $Z \subset X$ forms a cycle, then $\tau^n(Z) = Z$. It should be remarked that the set functions τ_i first appeared in Sternfeld [20], where instead of the distance functions $d_i(\mathbf{x})$ general continuous functions are involved.

The following theorem is valid.

Theorem 3.1 *Let X be a compact subset of $\mathbb{R}^d$. If $\cap_{n=1,2,\ldots}\tau^n(X) = \emptyset$, then the set $\mathcal{R}(S)$ is dense in $C(X)$.*

Since functions of the form $g_i\ (\|\mathbf{x} - \mathbf{c}_i\|)$ generate a subalgebra of the space $C(X)$, Theorem 3.1 immediately follows from a general result of Sproston and Straus [19] proved for a sum of continuous function algebras.

The following theorem establishes a sufficient condition and also a necessary condition for the density of RBF neural networks $\mathcal{G}(g, S)$ in $C(X)$.

Theorem 3.2 *Assume g is a continuous p-th degree ($1 \leq p < \infty$) integrable function, or g is a nonconstant continuous, bounded function, which has a limit at infinity (or minus infinity). Then the following statements hold:*

(a) If $\bigcap_{n=1,2,\ldots} \tau^n(X) = \emptyset$, then the set $\mathcal{G}(g, S)$ is dense in $C(X)$;

(b) If $\mathcal{G}(g, S)$ is dense in $C(X)$, then the set X does not contain cycles (with respect to S).

Proof (a) Suppose $\bigcap_{n=1,2,\ldots} \tau^n(X) = \emptyset$. By Theorem 3.1, the set $\mathcal{R}(S)$ is dense in $C(X)$. Hence for any function $f \in C(X)$ and any positive real ε there exist continuous functions g_i, $i = 1, ..., k$ such that

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} g_i\ (\|\mathbf{x} - \mathbf{c}_i\|) \right| < \frac{\varepsilon}{k+1} \tag{3.2}$$

for all $\mathbf{x} \in X$. Since X is compact, the sets $Y_i = \{\|\mathbf{x} - \mathbf{c}_i\| :\ \mathbf{x} \in X\}$, $i = 1, 2, ..., k$ are compacts as well. We know from Section 2 that the function g is not mean-periodic and hence the set

$$span\{g(t - \theta) :\ \theta \in \mathbb{R}\}$$

is dense in $C(\mathbb{R})$ in the topology of uniform convergence on compacta (see the proofs of Theorems 2.1 and 2.2). It follows that for the above ε there exist $c_{ij}, \theta_{ij} \in \mathbb{R}$, $i = 1, 2, ..., k$, $j = 1, ..., m_i$ such that

$$\left| g_i(t) - \sum_{j=1}^{m_i} c_{ij} g(t - \theta_{ij}) \right| < \frac{\varepsilon}{k+1} \tag{3.3}$$

for all $t \in Y_i,\ i = 1, 2, ..., k$. From (3.2) and (3.3) we obtain that

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} \sum_{j=1}^{m_i} c_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_{ij}) \right| < \varepsilon. \tag{3.4}$$

for all $\mathbf{x} \in X$. Thus $\overline{\mathcal{G}(g, S)} = C(X)$.

(b) Suppose $\mathcal{G}(g, S)$ is dense in $C(X)$. Then for an arbitrary positive real number ε, inequality (3.4) holds with some coefficients $c_{ij}, \theta_{ij},\ i = 1, 2, ..., k,\ j = 1, ..., m_i$. Since for each $i = 1, 2, ..., k$, the function $\sum_{j=1}^{m_i} c_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_{ij})$ is a function of the form $g_i(\|\mathbf{x} - \mathbf{c}_i\|)$, it follows from (3.4) that the subspace $\mathcal{R}(S)$ is dense in $C(X)$. Let us prove that X does not contain a cycle. Assume the contrary. Assume X contains a cycle, which we denote by $l = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be the vector known from Definition 3.1. Introduce the functional

$$F : C(X) \to \mathbb{R},\ F(f) = \sum_{j=1}^{n} \lambda_j f(\mathbf{x}_j).$$

Clearly, F is a linear bounded functional with the norm $\sum_{j=1}^{n} |\lambda_j|$. It is an exercise to check that $F(h) = 0$ for any $h \in \mathcal{R}(S)$. By Urysohn's lemma, there exists a continuous function $f_0 : X \to \mathbb{R}$ such that $f_0(\mathbf{x}_j) = 1$ if $\lambda_j > 0$, $f_0(\mathbf{x}_j) = -1$ if $\lambda_j < 0$ and $-1 < f_0(\mathbf{x}) < 1$, for any $\mathbf{x} \in X \backslash l$. For this function, $F(f_0) \neq 0$. We have constructed a nonzero annihilating functional F. The existence of such a functional means that $\mathcal{R}(S)$ cannot be dense in $C(X)$. The obtained contradiction proves the 2nd statement of the theorem.

At the end we want to point out that the solution to the density problem for RBF neural networks with only two fixed centroids are geometrically explicit. In this special case, we can completely characterize all compact sets $X \subset \mathbb{R}^d$ for which $\mathcal{G}(g, S)$ is dense in $C(X)$. To formulate the theorem, consider the following relation between points in X. The relation $\mathbf{x} \sim \mathbf{y}$ when $\mathbf{x}$ and $\mathbf{y}$ belong to some path in X defines an equivalence relation. The equivalence classes are called orbits (see [13]).

The following theorem holds.

Theorem 3.3 *Assume g is a continuous p-th degree ($1 \leq p < \infty$) integrable function, or g is a nonconstant continuous, bounded function, which has a limit at infinity (or minus infinity). Assume X is a compact subset of $\mathbb{R}^d$ with all its orbits closed and $S = \{\mathbf{c}_1, \mathbf{c}_2\}$ is the set of fixed centroids. Then the set $\mathcal{G}(g, S)$ is dense in $C(X)$ if and only if X contains no closed paths.*

The proof can be carried out in a similar way to the one given for the previous theorem. Only instead of Theorem 3.1 we use the following result, which is a corollary of the general result of Marshall and O'Farrell [13] on the uniform approximation by a sum of two function algebras: If all orbits of X are closed, then for the density of $\mathcal{R}(S)$ in $C(X)$ it is necessary and sufficient that X contain no closed paths.

Remark By definition, a closed path is a trace of some point jumping from one position to another, alternatively on the spheres $\|\mathbf{x} - \mathbf{c}_1\| = r_1$, $\|\mathbf{x} - \mathbf{c}_2\| = r_2$ (r_1 and r_2 are not fixed), and at the end returning to its primary position. In $\mathbb{R}^2$ the circles $\|\mathbf{x} - \mathbf{c}_1\| = r_1$, $\|\mathbf{x} - \mathbf{c}_2\| = r_2$ form a circular grid. Thus for density $G(g, S)$ in $C(X)$ it is necessary and sufficient that X does not contain any sequence of vertices (intersection points) $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n, \mathbf{x}_1\}$ of this grid with the premise that the pairs $\mathbf{x}_i, \mathbf{x}_{i+1}$ and $\mathbf{x}_{i+1}, \mathbf{x}_{i+2}$ lie on different circles.

Acknowledgements The authors would like to thank the anonymous reviewers for their insightful comments and suggestions, which helped to improve the quality of the paper.

Author Contributions Aysu Ismayilova wrote Section 2. Muhammad Ismayilov wrote Section 3. Both authors wrote Section 1 and reviewed the manuscript.

Funding statement No funding was received for this study.

Data Availibility Statement Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Information on conflict of interest The authors declare that they have no conflicts of interest.

Compliance with ethical standards This article does not contain any studies involving animals performed by any of the authors. This article does not contain any studies involving human participants performed by any of the authors.

Competing interest We declare that we have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Agarwal, V., Bhanot, S.: Radial basis function neural network-based face recognition using firefly algorithm. Neural Comput. Appl. **30**(8), 2643–2660 (2018)
2. Bishop, C.: Improving the generalisation properties of radial basis function neural networks. Neural Comput. **3**(4), 579–588 (1991)
3. Broomhead, D.S., Lowe, D.: Multivariable function interpolation and adaptive networks. Complex Systems **2**, 321–355 (1988)
4. Fath, A.H., Madanifar, F., Abbasi, M.: Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. Petroleum **6**(1), 80–91 (2020)
5. Funahashi, K.: On the approximate realization of continuous mappings by neural networks. Neural Netw. **2**, 183–192 (1989)
6. Huang, G.B., Chen, L., Siew, C.K.: Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Networks **17**(4), 879–892 (2006)
7. Ismailov, V. E.: Ridge functions and applications in neural networks. Mathematical Surveys and Monographs, 263. American Mathematical Society, Providence, RI,186. (2021)
8. Karamichailidou, D., Kaloutsa, V., Alexandridis, A.: Wind turbine power curve modeling using radial basis function neural networks and tabu search. Renewable Energy **163**, 2137–2152 (2021)
9. Khan, S., Naseem, I., Malik, M.A., Togneri, R., Bennamoun, M.: A fractional gradient descent-based RBF neural networks. Circuits, Systems, Signal Processing. **37**, 5311–5332 (2018)
10. Li, X., Sun, Y.: Application of RBF neural network optimal segmentation algorithm in credit rating, Neural Computing and Applications. 8227–8235. (2021)
11. Liao, Y., Fang, S.C., Nuttle, H.L.W.: Relaxed conditions for radial-basis function networks to be universal approximators. Neural Netw. **16**(7), 1019–1028 (2003)
12. Lippman, R.P.: Pattern classification using neural networks. IEEE Commun. Mag. **27**, 47–64 (1989)
13. Marshall, D.E., O'Farrell, A.G.: Uniform approximation by real functions. Fund. Math. **104**, 203–211 (1979)
14. Moody, J., Darken, C.: Learning with localized receptive fields. In: Proceedings of the 1988 Connectionist Models Summer School, Morgan-Kaufmann, Publishers, 1988
15. Park, J., Sanberg, I.W.: Approximation and radial-basisfunction networks. Neural Comput. **5**, 305–316 (1993)
16. Park, J., Sanberg, I.W.: Universal approximation using radial-basis-function networks. Neural Comput. **2**, 246–257 (1991)
17. Pinkus, A.: Approximation theory of the MLP model in neural networks. Acta Numer **8**, 143–195 (1999)

18. Schwartz, L.: Theorie generale des fonctions moyenne-periodiques. Ann. Math. **48**, 857–928 (1947)
19. Sproston, J.P., Strauss, D.: Sums of subalgebras of $C(X)$. J. London Math. Soc. **45**, 265–278 (1992)
20. Sternfeld, Y.: Uniformly separating families of functions. Israel J. Math. **29**, 61–91 (1978)
21. Vitushkin, A.G., Henkin, G.M.: Linear superpositions of functions. (Russian), Uspehi Mat. Nauk. **22**, 77–124. (1967)
22. Wang, H., Liu, K., Wu, Y., Wang, S., Zhang, Z., Li, F., Yao, J.: Image reconstruction for electrical impedance tomography using radial basis function neural network based on hybrid particle swarm optimization algorithm. IEEE Sens. J. **21**(2), 1926–1934 (2020)
23. Wang, R., Li, D., Miao, K.: Optimized radial basis function neural network based intelligent control algorithm of unmanned surface vehicles. Journal of Marine Science and Engineering. **8**(3), 210 (2020)
24. Wu, Y., Wang, H., Zhang, B., Du, K.L.: Using radial basis function networks for function approximation and classification. ISRN Appl. Math **2012**, 1–34 (2012)