JID: NEUCOM [m5G;August 23, 2018;22:55]

Neurocomputing 000 (2018) 1–8

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Approximation capability of two hidden layer feedforward neural networks with fixed weights

Namig J. Guliyev, Vugar E. Ismailov*

Institute of Mathematics and Mechanics, Azerbaijan National Academy of Sciences, 9 B. Vahabzadeh str., Baku AZ1141, Azerbaijan

ARTICLE INFO

Article history:
Received 31 August 2017
Revised 16 June 2018
Accepted 17 July 2018
Available online xxx

Communicated by Long Cheng

2010 MSC:
41A30
41A63
65D15
68T05
92B20

Keywords:
Multilayer feedforward neural network
Hidden layer
Sigmoidal function
Activation function
Weight
The Kolmogorov superposition theorem

ABSTRACT

We algorithmically construct a two hidden layer feedforward neural network (TLFN) model with the weights fixed as the unit coordinate vectors of the d-dimensional Euclidean space and having $3d+2$ number of hidden neurons in total, which can approximate any continuous d-variable function with an arbitrary precision. This result, in particular, shows an advantage of the TLFN model over the single hidden layer feedforward neural network (SLFN) model, since SLFNs with fixed weights do not have the capability of approximating multivariate functions.

1. Introduction

The topic of artificial neural networks is an important and vibrant area of research in modern science. This is due to a large number of application areas. Nowadays, neural networks are being successfully applied in areas as diverse as computer science, finance, medicine, geology, engineering, physics, etc. Perhaps the greatest advantage of neural networks is their ability to be used as an arbitrary function approximation mechanism. In this paper, we are interested in questions of density (or approximation with arbitrary accuracy) of the multilayer feedforward neural network (MLFN) model. Approximation capabilities of this model have been well studied for the past 30 years. Choosing various activation functions σ it was shown in a great number of papers that MLFNs can approximate any continuous function with an arbitrary precision. The most simple MLFN model is the single hidden layer feedforward neural network (SLFN) model. This model evaluates a multivariate function

$$\sum_{i=1}^{k} c_i \sigma(\mathbf{w}^i \cdot \mathbf{x} - \theta_i) \tag{1.1}$$

of the variable $\mathbf{x} = (x_1, \dots, x_d)$, $d \geq 1$. Here the weights $\mathbf{w}^i$ are vectors in $\mathbb{R}^d$, the thresholds θ_i and the coefficients c_i are real numbers, and the activation function σ is a univariate function. A multiple hidden layer network is defined by iterations of the SLFN model. For example, the output of the two hidden layer feedforward neural network (TLFN) model with k units in the first layer, m units in the second layer and the input $\mathbf{x} = (x_1, \dots, x_d)$ is

$$\sum_{i=1}^{m} e_i \sigma\left(\sum_{j=1}^{k} c_{ij} \sigma(\mathbf{w}^{ij} \cdot \mathbf{x} - \theta_{ij}) - \zeta_i \right).$$

Here d_i, c_{ij}, θ_{ij} and γ_i are real numbers, $\mathbf{w}^{ij}$ are vectors of $\mathbb{R}^d$, and σ is a fixed univariate function.

In many applications, it is convenient to take an activation function σ as a *sigmoidal function*, which is defined as

$$\lim_{t \to -\infty} \sigma(t) = 0 \quad \text{and} \quad \lim_{t \to +\infty} \sigma(t) = 1.$$

* Corresponding author.
E-mail address: vugaris@mail.ru (V.E. Ismailov).

https://doi.org/10.1016/j.neucom.2018.07.075
0925-2312/

The literature on neural networks abounds with the use of such functions and their superpositions (see, e.g., [1,3,5,7–11,13,15–17,20,27,28,33,40,42]).

The possibility of approximating a continuous function on a compact subset of $\mathbb{R}^d$, $d \geq 1$, by SLFNs with a sigmoidal activation function has been tremendously studied in many papers. To the best of our knowledge, Gallant and White [16] were the first to prove the universal approximation property for the SLFN model with a sigmoidal activation function. Their activation function, called the *cosine squasher*, has the ability to generate any trigonometric series. As such, this function has the density property. Carroll and Dickinson [4] implemented the inverse Radon transformation to approximate L^2 functions, using any continuous sigmoidal function as an activation function. Cybenko [13] proved that SLFNs with a continuous sigmoidal activation function can approximate any continuous function with arbitrary accuracy on compact subsets of $\mathbb{R}^d$. Funahashi [15], independently of Cybenko, proved the density property for a continuous monotone sigmoidal function. Hornik, Stinchcombe and White [22] proved density of SLFNs with a discontinuous bounded sigmoidal function. Kůrková [33] showed that staircase-like functions of any sigmoidal type has the capability of approximating continuous univariate functions on any compact subset of $\mathbb{R}$ within arbitrarily small tolerance. This result was substantially used in Kůrková's further results, which showed that a continuous multivariate function can be approximated arbitrarily well by TLFNs with a sigmoidal activation function (see [32,33]). Chen et al. [6] generalized the result of Cybenko by proving that any continuous function on a compact subset of $\mathbb{R}^d$ can be approximated by SLFNs with a bounded (not necessarily continuous) sigmoidal activation function. Almost the same result was independently obtained by Jones [30]. Costarelli and Spigler [9] constructed special sums of the form (1.1), using a given function $f \in C[a, b]$. They then proved that these sums approximate f within any degree of accuracy. In their result, similar to [6], σ is any bounded sigmoidal function. Chui and Li [7] proved that SLFNs with a continuous sigmoidal activation function having integer weights and thresholds can approximate continuous univariate functions on any compact subset of the real line.

In a number of subsequent papers, which considered the density problem for the SLFN model, nonsigmoidal activation functions were allowed. Here we cite a few of them. The papers by Stinchcombe and White [46], Cotter [12], Hornik [21], Mhaskar and Micchelli [42] are among many others. It should be remarked that the more general result in this direction belongs to Leshno et al. [34]. They proved that the necessary and sufficient condition for any continuous activation function to have the density property is that it not be a polynomial. For more detailed discussion of the density problem, see the review paper by Pinkus [43].

The above results show that SLFNs with various activation functions enjoy the universal approximation property. In recent years, the theory of neural networks has been developed further in this direction. For example, from the point of view of practical applications, SLFNs with a restricted set of weights have gained a special interest (see, e.g., [14,23,25,26,29,35]). It was proved that SLFNs with some restricted set of weights still possess the universal approximation property. For example, Stinchcombe and White [46] showed that SLFNs with a polygonal, polynomial spline or analytic activation function and a bounded set of weights have the universal approximation property. Ito [27,28] investigated this property of networks using monotone sigmoidal functions, with only weights located on the unit sphere. In [23,25,26], the second coauthor considered SLFNs with weights varying on a restricted set of directions, and gave several necessary and sufficient conditions for good approximation by such networks. For a set of weights consisting of two directions, he showed that there is a geometrically explicit solution to the problem. Hahm and Hong [20] went further in this direction, and showed that SLFNs with fixed weights can approximate arbitrarily well any continuous univariate function. Since fixed weights reduce the computational expense and training time, this result is of particular interest. In a mathematical formulation, the result says that for a bounded measurable sigmoidal function σ, networks of the form $\sum_{i=1}^{k} c_i \sigma(\alpha x - \theta_i)$ are dense in $C[a, b]$. Cao and Xie [3] strengthened this result by specifying the number of hidden neurons to realize ε-approximation to any continuous function. By implementing modulus of continuity, they established Jackson-type upper bound estimations for the approximation error.

Approximation capabilities of SLFNs with fixed weights were also analyzed in Lin et al. [37]. Taking the activation function σ as a continuous, even and 2π-periodic function, the authors of [37] showed that neural networks of the form $\sum_{i=1}^{r} c_i \sigma(x - x_i)$ can approximate any continuous function on $[-\pi, \pi]$ with an arbitrary precision ε. Note that all the weights are fixed equal to 1, and consequently do not depend on ε. To prove this, they first gave an integral representation for trigonometric polynomials, and constructed explicitly a network with the weight 1 that approximates this integral representation. Finally, the obtained result for trigonometric polynomials was used to prove a Jackson-type upper bound for the approximation error.

Note that SLFNs with a fixed number of weights cannot approximate d-variable functions if $d > 1$. That is, if in (1.1) we have n different weights $\mathbf{w}^i$ (n is fixed), then there exist a compact set $Q \subset \mathbb{R}^d$ and a function $f \in C(Q)$, which cannot be approximated arbitrarily well by the networks formed as (1.1). This follows from a result of Lin and Pinkus on sums of n ridge functions (see [38, Theorem 5.1]). For details, see our recent paper [19]. Thus the above results of Hahm and Hong [20], Cao and Xie [3], Lin et al. [37] cannot be generalized to the d-dimensional case if one allows only the SLFN model of neural networks.

It should be remarked that in all of the above-mentioned works the number of neurons k in the hidden layer is not fixed. As such to achieve a desired precision one may take an excessive number of hidden neurons. Unfortunately, practicality decreases with the increase of the number of neurons in the hidden layer. In other words, SLFNs are not always effective if the number of neurons in the hidden layer is prescribed. More precisely, they are effective if and only if we consider univariate functions. In [18], we consider constructive approximation on any finite interval of $\mathbb{R}$ by SLFNs with a fixed number of hidden neurons. We construct algorithmically a smooth, sigmoidal, almost monotone activation function σ providing approximation to an arbitrary univariate continuous function within any degree of accuracy. Note that the result of [18] is not applicable to multivariate functions.

The first crucial step in investigating approximation capabilities of MLFNs with a prescribed number of hidden neurons was made by Maiorov and Pinkus [41]. Their remarkable result revealed that TLFNs with $3d$ units in the first layer and $6d + 3$ units in the second layer can approximate an arbitrary continuous d-variable function. Using a different activation function than in [41], the second coauthor [24] showed that the number of neurons in hidden layers can be reduced to d and $2d + 2$ respectively. Note that the results of both papers carry a theoretical character, as they indicate only the existence of the corresponding TLFNs, their activation functions.

We see that in each result above at least one of the following general properties is violated.

1. The number of hidden neurons is fixed.
2. The weights are fixed.
3. The activation function is computable.
4. The network has the capability of approximating d-variable functions in the case $d > 1$.