Interpolation of Generalized Functions Using Artificial Neural Networks

Raghu T. Mylavarapu[1], Bharadwaja Krishnadev Mylavarapu[2], Uday Shankar Sekhar[3]

[1]Regional Development Center, KPMG LLP, Montvale, NJ, USA
[2]Business Analytics Lead, Department of Analytics, PLM COE, PDM, John Deere Co, Moline, IL, USA
[3]Financial Serivces Office, Global Data Analytics Platform, Ey, New York, USA
Email: rmylavarapu@kpmg.com, Bharadwajakrishnadev@johndeere.com, Uday.Sekhar@Ey.Com

Abstract

In this paper we employ artificial neural networks for predictive approximation of generalized functions having crucial applications in different areas of science including mechanical and chemical engineering, signal processing, information transfer, telecommunications, finance, etc. Results of numerical analysis are discussed. It is shown that the known Gibb's phenomenon does not occur.

Keywords

Function Interpolation, Hidden Layer, Weight, Bias, Machine Learning, Distributions, Dirac's Function, Heaviside's Function, Gibb's Phenomenon

1. Introduction: Main Definitions, Representations and Relations between Known Distributions

Generalized functions are used in modeling and analysis of applied systems in various areas of science including engineering, finance, control theory, etc. Practically every object or phenomenon containing discontinuity, switches or localized character, in principle, can be described in terms of generalized functions or distributions. The theory of generalized functions is developed by Sergei Sobolev (1930s) and Laurent Schwartz (1940s) (for details see the monographs [1] [2] [3] and the references therein). Nevertheless, George Green and Oliver Heaviside used generalized functions in their research much earlier. The Dirac's δ function defined by

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ \infty, & x = 0, \end{cases} \tag{1}$$

is used in the Green's representation formula for the general solution of non-homogeneous boundary value problems. Later in 1930s, Paul Dirac systematically used the δ function to describe a point charge localized at a given point. In practical analysis, the definition of the Dirac's δ(1) must be supplemented by

$$\int_{-\infty}^{\infty} \delta(x)\,\mathrm{d}x = 1, \quad \int_{-\infty}^{\infty} f(x)\delta(x-x_0)\,\mathrm{d}x = f(x_0), \tag{2}$$

for arbitrary continuous function *f*.

On the other hand, Heaviside used the θ function given by

$$\theta(x) = \begin{cases} 1, & x > 0, \\ 0, & x < 0, \end{cases} \tag{3}$$

to extend the notion of the Laplace integral transform in telegraphic communications. At this, the value θ(0) depends on particular problem and can be one of 0, 1 or 0:5.

According to the definition, the Dirac's and Heaviside's functions are related by

$$\theta(x) = \int_{-\infty}^{x} \delta(\xi)\,\mathrm{d}\xi, \tag{4}$$

In other words, θ is the antiderivative of δ in the sense of generalized functions. On the other hand, the function max {*x*, 0} is the antiderivative of θ, therefore

$$\int_{-\infty}^{x} \theta(\xi)\,\mathrm{d}\xi = \int_{-\infty}^{x} (x-\xi)\delta(\xi)\,\mathrm{d}\xi = \max\{x,0\},$$

which direct follows from the second equality in (2) with *f*(*x*) = *x*.

There exists a linear relation between the Heaviside's generalized function and the sign function defined by:

$$\theta(x) = \frac{1}{2} + \frac{1}{2}\sin(x).$$

Evidently, here the value θ(0) = 0.5 is considered. However, it is possible to write such a formula with θ(0) = 0 or θ(0) = 1.

Other known generalized functions can be defined through θ. For instance, the characteristic function defined by

$$X_{[a,b]}(x) = \begin{cases} 1, & x \in [a,b], \\ 0, & \text{else}, \end{cases}$$

can be expressed in terms of θ according to

$$X_{[a,b]} = \theta(x-a) - \theta(x-b).$$

The rectangular function defined by

$$rect(x) = \begin{cases} 1, & |x| < 0.5, \\ 0.5, & x = 0.5, \\ 0, & |x| > 0.5, \end{cases}$$

can be expressed in terms of θ as follows:

$$rect(x) = \theta\left(x + \frac{1}{2}\right) - \theta\left(x - \frac{1}{2}\right)$$

or

$$rect(x) = \theta\left(\frac{1}{2} - x^2\right).$$

As above, here the value $\theta(0) = 0.5$ is considered as well.

Another well-known generalized function is defined through the convolution

$$R(x) = \theta(x) * \theta(x),$$

where $*$ denotes the convolution operation. This function is called ramp function and has many applications in engineering (it is used in the so-called half-wave rectification, which is used to convert alternating current into direct current by allowing only positive voltages), artificial neural networks (it serves as an activation function), finance, statistics, fluid mechanics, etc.

According to the definition of the Heaviside function, the rump function can be represented also as

$$R(x) = \theta(x) * \theta(x) = \int_{-\infty}^{x} \theta(x - \xi)\theta(\xi)\,d\xi = \int_{-\infty}^{x} \theta(\xi)\,d\xi = x\theta(x)$$

2. Approximation of Main Generalized Functions by Means of Locally Measurable Functions

The theory of generalized functions is a very well developed mathematics subject crucial for rigorous analysis of many applied systems. Nevertheless, their rigorous definitions are completely useful in numerical analysis, because they are even not proper functions. In numerical analysis proper functional approximations of the generalized functions are used instant.

In practice, the approximation of generalized functions is based on construction of a sequence f_n of measurable functions giving the desired generalized function in limit when $n \to \infty$. For instance, the sequence

$$\delta_n(x) = \frac{n}{\sqrt{\pi}} \exp\left(-n^2 x^2\right),$$

which is also called Gauss kernel, tend to the Dirac's generalized function when $n \to \infty$. The sequence δ_n is called δ-like sequence. Several other δ-like sequences can be found in literature. Examples include

$$\delta_n(x) = \frac{1}{\pi} \frac{n}{1 + n^2 x^2},$$

which is also called Poisson kernel

$$\delta_n(x) = \frac{1}{2\pi s(x)} \sin\left(nx + \frac{1}{2x}\right), \quad s(x) = \sin\left(\frac{1}{2}x\right),$$

which is also called Dirichlet kernel, and

$$\delta_n(x) = \frac{1}{2\pi n s^2(x)} \sin^2\left(\frac{n}{2}x\right),$$

which is also called Fejér kernel. Note that for all mentioned kernels,

$$\delta_n \in L^1_{loc}(-\infty, \infty)$$

i.e. they are locally measurable functions.

Taking into account (4), similar θ-like sequences can be constructed for approximating Heaviside's θ. For instance,

$$\theta_n(x) = \frac{1}{2}\left[1 + \tanh(nx)\right], \tag{5}$$

often referred to as logistic function,

$$\theta_n(x) = \frac{1}{2}\left[1 + \frac{2}{\pi}\arctan(nx)\right],$$

$$\theta_n(x) = \frac{1}{2}\left[1 + erf(nx)\right],$$

also called erf-approximation,

$$\theta_n(x) = \frac{1}{2}\left[1 + \frac{2}{\pi}si(\pi nx)\right],$$

$$\theta_n(x) = \exp\left[-\exp(-nx)\right],$$

Similar sequences can be constructed for the functions sign, χ, rect and R above. For example,

$$rect_n(x) = \exp\left[-\exp\left(-n\left(\frac{1}{4} - x^2\right)\right)\right]$$

and

$$rect_n(x) = \frac{1}{2}\left[1 + \tanh\left(n\left(\frac{1}{4} - x^2\right)\right)\right],$$

can be viewed as rect-like sequences.

On the other hand, the expression

$$R_n = x\exp\left[-\exp(-nx)\right]$$

can be used as approximation to the ramp function.

3. Approximation of Generalized Functions Using Artificial Neural Networks

Generalized this paper we show through numerical experiments that artificial neural networks can provide a very fast and efficient approximation for generalized functions using any of the approximate formula above. There is a huge body of references devoted to the theory and implementation of artificial neural networks for approximation of functions. We refer to [4] [5] [6] [7] [8] and the references therein.

The neural network providing approximation consists of an input layer, a hidden layer and output layer. The quadratic error of approximation

$$\varepsilon = \left(f\left(x \right) - f_{app}\left(x \right) \right)^2$$

is considered, where f is the original function and f_{app} is its approximation. Moreover, in all examples below the θ-like sequence (5) considered. Other sequences can be applied exactly in the same way. The learning rate is always fixed to 10^{-3} for simplicity.

Approximation of the rect function for different number of nodes is presented in **Figure 1**. A better approximation with less error can be obtained by increasing the learning rate or the number of nodes. The error is plotted on **Figure 2**, from which it is obvious that the least error is $\varepsilon \sim 10^{-4}$ It is evident from **Figure 3** that the known Gibbs phenomenon does not occur here [9].
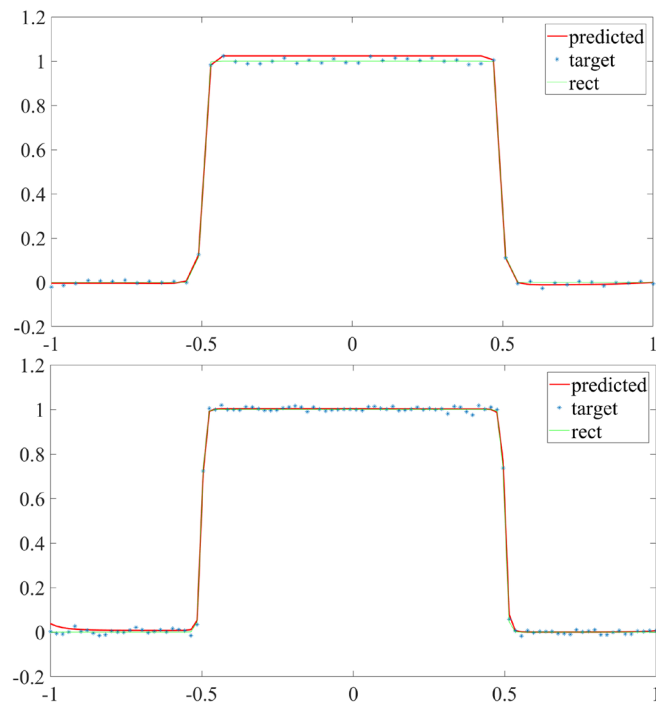

Figure 1. Approximation of rect function in [−1, 1] with 50 nodes (upper) and 100 nodes (lower).
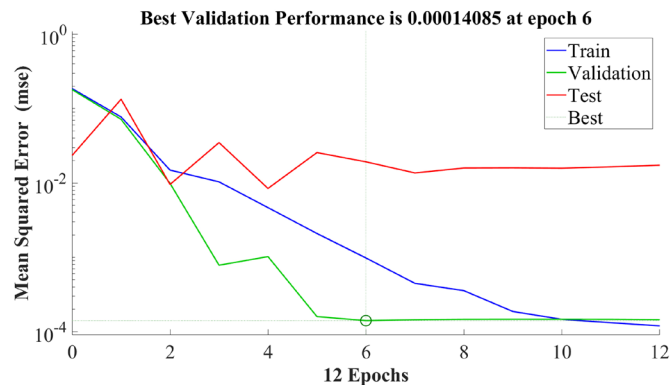

Figure 2. Quadratic error approximation of rect function in [−1, 1] with 100 nodes.
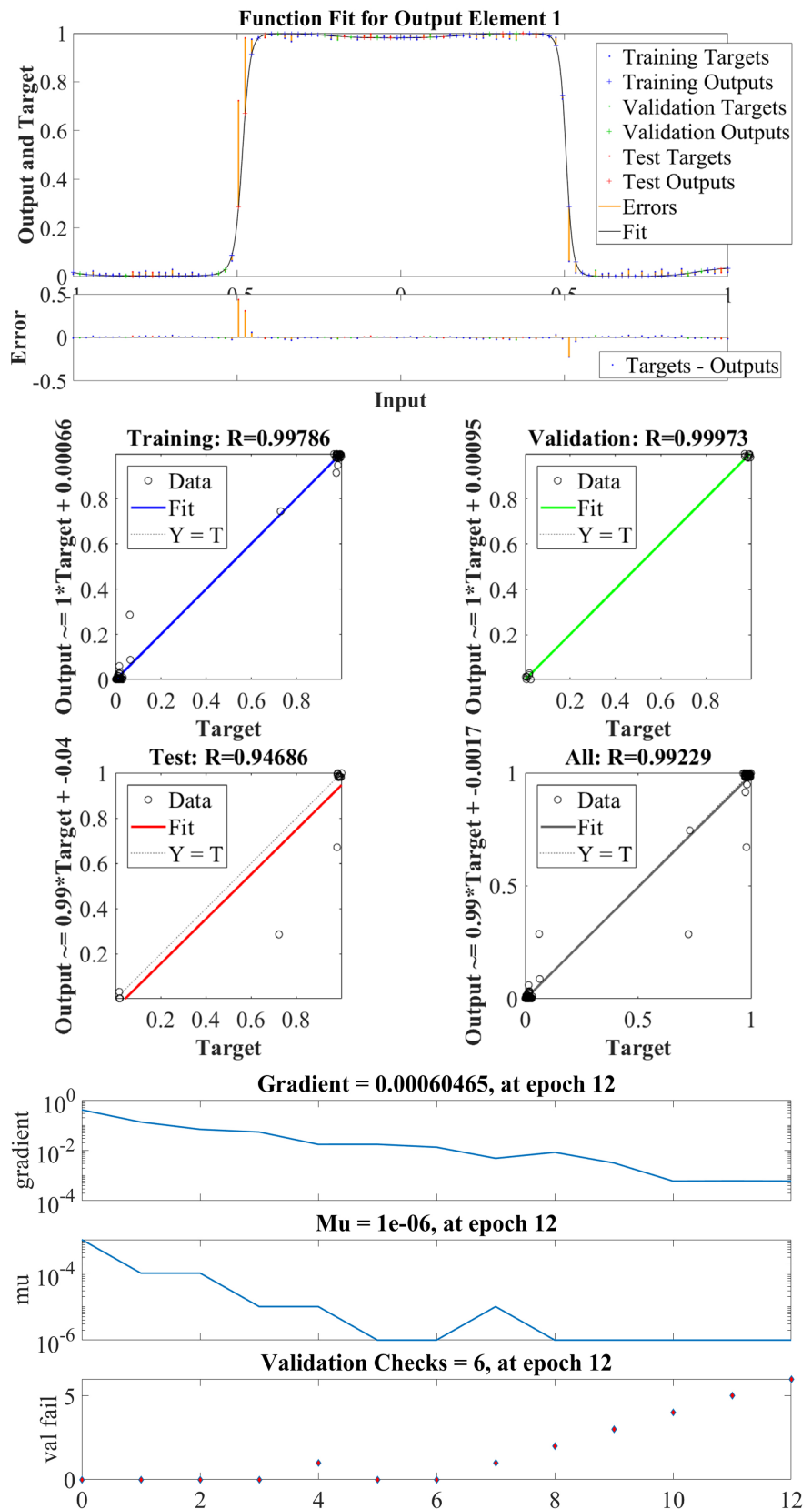
Figure 3. Function fit (upper) and regression behavior (middle) and network performance (lower) for rect function in [−1, 1] with 100 nodes.

4. Conclusions

Possibilities if artificial neural network approximation of generalized functions is considered by means of locally measurable approximations of the Dirac's delta function and Heaviside's theta function. Considering the quadratic error of approximation, the rect function is approximated taking into account the relation between the rect and Heaviside's theta functions. It is shown that due to the usage of neural networks, the Gibbs phenomenon does not occur. Using similar representation formulas and other approximations to the Heaviside's function, the characteristic or sign functions can also be approximated by artificial neural networks.

The results can be employed in numerical analysis of problems containing discontinuous phenomena, switching dynamics, etc.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Teodorescu, P.P., Kecs, W.W. and Toma, A. (2003) Distribution Theory: With Applications in Engineering and Physics. Wiley-VCH Verlag, Weinheim.

[2] Vladimirov, V.S. (2002) Methods of the Theory of Generalized Functions. Taylor & Francis, London, New York.

[3] Grubb, G. (2009) Distributions and Operators. Springer, Berlin, Heidelberg.

[4] Weigend, A.S. and Gershenfeld, N.A. (Eds.) (1994) Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley, Reading.

[5] Zhang, W. and Barrion, A. (2006) Function Approximation and Documentation of Sampling Data Using Artificial Neural Networks. *Environmental Monitoring and Assessment*, **122**, 185. https://doi.org/10.1007/s10661-005-9173-6

[6] Zainuddin, Z. and Ong, P. (2007) Function Approximation Using Artificial Neural Networks. *WSEAS Transactions on Mathematics*, **1**, 173-178.

[7] Zhang, Q. and Benveniste, A. (1991) Approximation by Nonlinear Wavelet Networks. *IEEE Transactions on Neural Networks*, **6**, 3417-3420.

[8] Ferrari, S. and Stengel, F. Smooth Function Approximation Using Neural Networks. Lecture Notes. http://ce.sharif.edu/courses/84-85/2/ce667/resources/root/Seminar_no_10/tnnlatestmanu.pdf

[9] Weisstein, E.W. (2013) Gibbs Phenomenon. From MathWorld—A Wolfram Web Resource.