On Some Families of Recurrence Generated Activation Functions

Vesselin Kyurkchiev[1], Anton Iliev[1,2] and Nikolay Kyurkchiev[2]
[1] Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski,
24, Tzar Asen Str., 4000 Plovdiv, Bulgaria, e–mails: vkyurkchiev@gmail.com, aii@uni–plovdiv.bg
[2] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria, e–mail: nkyurk@math.bas.bg

Abstract

In this note we construct some families of recurrence generated activation functions based on Cauchy activation function (CAF) and Hyperbolic–secant activation function (HSAF).

We prove estimates for the Hausdorff approximation of the Heaviside function $h(t)$ by means of these families. Numerical examples, illustrating our results are given.

Keywords: *Recurrence Generated Cauchy Activation Functions (CAF), Recurrence Generated Hyperbolic–secant Activation Functions (HSAF), Heaviside Function, Hausdorff Distance, Upper and Lower Bounds.*

1. Introduction

Sigmoidal functions (also known as “activation functions”) find multiple applications in many scientific fields, including biology, ecology, population dynamics, chemistry, demography, economics, geoscience, mathematical psychology, probability, sociology, political science, financial mathematics, statistics, fuzzy set theory, insurance mathematics, neural networks, to name a few [1]–[19].

We study the distance between the Heaviside function and a special class of sigmoidal functions, so–called activation functions.

The distance is measured in Hausdorff sense, which is natural in a situation when a Heaviside function is involved. Estimates for the Hausdorff distance are reported.

Constructive approximation by superposition of sigmoidal functions and the relation with neural networks and radial basis functions approximations is discussed in [19].

Any neural net element computes a linear combination of its input signals, and uses a logistic function to produce the result; often called “activation” function [20]– [21].

2. Preliminaries

Definition 1. The Cauchy activation function is defined for $b > 0$ by [22]:

$$f(t) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{t}{b}\right). \qquad (1)$$

for which $f(0) = \frac{1}{2}$.

Definition 2. The Hyperbolic–secant activation function is defined for $b > 0$ by [23]:

$$g(t) = \frac{2}{\pi}\arctan\left(e^{\frac{t}{b}}\right) \qquad (2)$$

for which $g(0) = \frac{1}{2}$.

Definition 3. The Heaviside step function $h(t)$ is defined by

$$h(t) = \begin{cases} 0, & \text{if} \quad t < 0, \\ [0,1], & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0. \end{cases} \qquad (3)$$

Definition 4. The Hausdorff distance (H–distance) $\rho(f,g)$ between two interval functions f, g on $\Omega \subseteq \mathrm{R}$, is the distance between their completed graphs $F(f)$ and $F(g)$ considered as closed subsets of $\Omega \times \mathrm{R}$ [24], [25]. More precisely,

$$\rho(f,g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} \|A - B\|,$$
$$\sup_{B \in F(g)} \inf_{A \in F(f)} \|A - B\|\},$$

wherein $\|.\|$ is any norm in R^2, e. g. the maximum norm $\|(t,x)\| = \max\{|t|,|x|\}$; hence the distance between the points $A = (t_A, x_A)$, $B = (t_B, x_B)$ in R^2 is $\|A - B\| = max(|t_A - t_B|, |x_A - x_B|)$.

3. Main Results

In this Section we construct some families of recurrence generated activation functions based on $f(t)$ and $g(t)$.

We prove estimates for the Hausdorff approximation of the Heaviside function $h(t)$ by means of these families.

3.1. The Family of Recurrence Generated Cauchy Activation Functions (CAF)

We consider the following family of recurrence generated activation functions:

$$f_{i+1}(t)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{1}{b}\left(t-\frac{1}{2}+f_i(t)\right)\right),\qquad (4)$$
$$i=0,1,2,\ldots;\ b\geq 0,$$

with

$$f_0(t)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{t}{b}\right);\ \ f_0(0)=\frac{1}{2}.\qquad (5)$$

Evidently, $f_{i+1}(0)=\frac{1}{2}$ for $i=0,1,2,\ldots,$.

Denote the number of recurrences by p.

Let us point out that the Hausdorff distance is a natural measuring criteria for the approximation of bounded discontinuous functions [17].

3.2. Approximation Issues

We study the Hausdorff distance d between the Heaviside function $h(t)$ and the family of (CAF)–functions (4)–(5).

Special case. Let $p=2$.

The H –distance $d_2(h(t), f_2(t))$ between the function $h(t)$ and the function $f_2(t)$ satisfies the relation:

$$f_2(d_2)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{1}{b}\left(d_2-\frac{1}{2}+f_1(d_2)\right)\right)\qquad (6)$$
$$=1-d_2.$$

The following Theorem gives upper and lower bounds for d_2

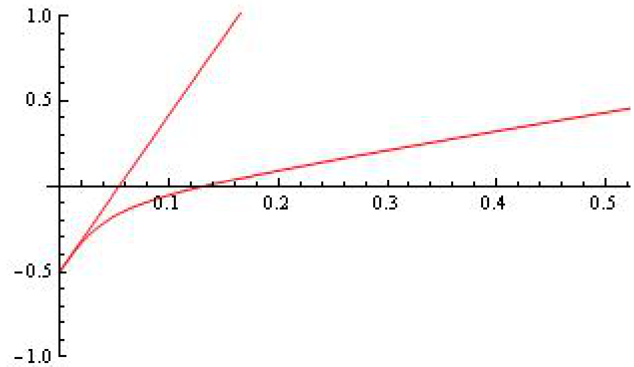

Fig. 1. The functions $F_2(d_2)$ and $G_2(d_2)$ for $b=0.2$.

Theorem 3.1. *For the Hausdorff distance d_2 between the function $h(t)$ and the function $f_2(t)$ the following inequalities hold:*

$$d_{l_2}=\frac{1}{\frac{5}{2}(1+\frac{1}{b\pi}+\frac{1}{(b\pi)^2}+\frac{1}{(b\pi)^3})}<d_2$$
$$<\frac{\ln\left(\frac{5}{2}(1+\frac{1}{b\pi}+\frac{1}{(b\pi)^2}+\frac{1}{(b\pi)^3})\right)}{\frac{5}{2}(1+\frac{1}{b\pi}+\frac{1}{(b\pi)^2}+\frac{1}{(b\pi)^3})}=d_{r_2}\qquad (7)$$

for $b<\frac{1}{\pi x}$, where x is the unique positive root of the polynomial equation

$$x^3+x^2+x+1-A=0;\ \ A=\frac{2}{5}e^{\frac{5}{4}}.\qquad (8)$$

Proof. We define the functions

$$F_2(d_2)=\frac{1}{2}+$$
$$\frac{1}{\pi}\arctan\left(\frac{1}{b}\left(d_2-\frac{1}{2}+f_1(d_2)\right)\right)-1+d_2\qquad (9)$$

and

$$G_2(d_2)=-\frac{1}{2}+\left(1+\frac{1}{b\pi}+\frac{1}{(b\pi)^2}+\frac{1}{(b\pi)^3}\right)d_2.\qquad (10)$$

From Taylor expansion

$$F_2(d_2)-G_2(d_2)=O(d_2^2)$$

we see that $G_2(d_2)$ approximates $F_2(d_2)$ with $d_2\to 0$ as $O(d_2^2)$ (cf. Fig. 1).

In addition $G_2'(d_2) > 0$ and for $\beta \geq 5$

$$G_2(d_{l_2}) < 0;\ G_2(d_{r_2}) > 0.$$

This completes the proof of the inequalities (7).

Remarks. For $A = \frac{2}{5}e^{\frac{5}{4}} \approx 1.39614$ we have $1 - A \approx -0.39614$ and from Descartes' rule of signs the algebraic equation (8) has unique positive root $x \approx 0.288722$. Then $b < 1.10248$.

Precise upper and lower bounds for the unique positive root of algebraic equation can be found in [26], [27].

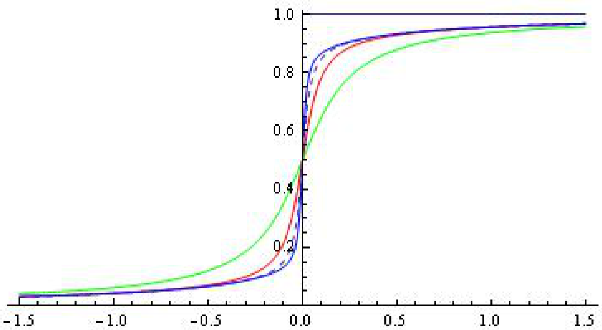

Fig. 2. Approximation of the $h(t)$ by (CAF)–functions for $b = 0.2$; The graphics of recurrence generated functions: f_0 (green), f_1 (red), f_2 (dashed) and f_3 (blue); Hausdorff distance: $d_0 = 0.228713$, $d_1 = 0.157674$, $d_2 = 0.1322$, $d_3 = 0.123957$.

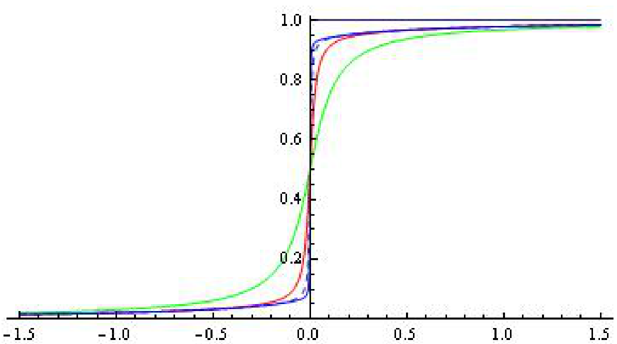

Fig. 3. Approximation of the $h(t)$ by (CAF)–functions for $b = 0.1$; The graphics of recurrence generated functions: f_0 (green), f_1 (red), f_2 (dashed) and f_3 (blue); Hausdorff distance: $d_0 = 0.16959$, $d_1 = 0.0931563$, $d_2 = 0.0691582$, $d_3 = 0.063748$.

The recurrence generated (CAF)–functions $f_0(t)$, $f_1(t)$, $f_2(t)$ and $f_3(t)$ for various b are visualized on Fig. 2–Fig. 3.

3.3. The Family of Recurrence Generated Hyperbolic–secant Activation Functions (HSAF)

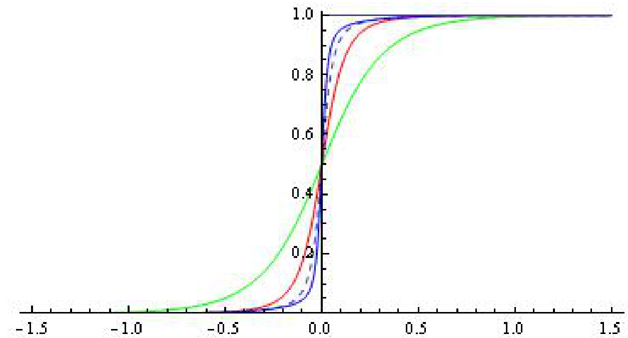

Fig. 4. Approximation of the $h(t)$ by (HSAF)–functions for $b = 0.2$; The graphics of recurrence generated functions: g_0 (green), g_1 (red), g_2 (dashed) and g_3 (blue); Hausdorff distance: $d_0 = 0.212162$, $d_1 = 0.127451$, $d_2 = 0.0893487$, $d_3 = 0.0692835$.

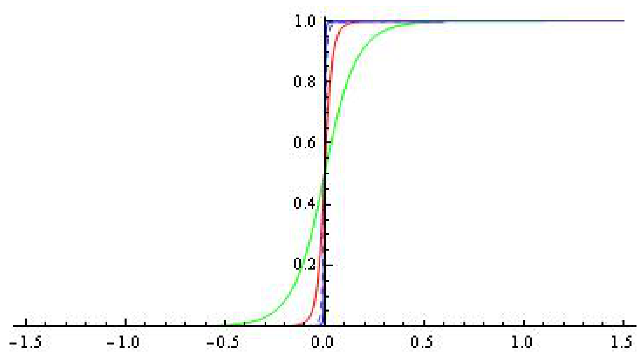

Fig. 5. Approximation of the $h(t)$ by (HSAF)–functions for $b = 0.1$; The graphics of recurrence generated functions: g_0 (green), g_1 (red), g_2 (dashed) and g_3 (blue); Hausdorff distance: $d_0 = 0.145697$, $d_1 = 0.0591496$, $d_2 = 0.026194$, $d_3 = 0.0122358$.

We consider the following family of recurrence generated activation functions:

$$g_{i+1}(t) = \frac{2}{\pi}\arctan\left(e^{\frac{1}{b}\left(t-\frac{1}{2}+g_i(t)\right)}\right), \quad (11)$$
$$i = 0,1,2,\ldots;\ b \geq 0,$$

with

$$g_0(t) = \frac{2}{\pi}\arctan\left(e^{\frac{t}{b}}\right);\ g_0(0) = \frac{1}{2}. \quad (12)$$

Evidently, $g_{i+1}(0) = \frac{1}{2}$ for $i = 0,1,2,\ldots,$.

The recurrence generated (HSAF)–functions $g_0(t)$, $g_1(t)$, $g_2(t)$ and $g_3(t)$ for various b are visualized on Fig. 4 – Fig. 5.

Based on the methodology proposed in the present note, the reader may formulate the corresponding approximation problems on his/her own.

4. Conclusion

A family of recurrence generated parametric activation functions is introduced finding application in neural network theory and practice.

Theoretical and numerical results on the approximation in Hausdorff sense of the Heaviside function by means of functions belonging to the family are reported in the paper.

We propose a software module within the programming environment *CAS Mathematica* for the analysis of the considered family of recurrence generated (CAF) and (HSAF) functions.

The module offers the following possibilities:

– generation of the activation functions under user defined values of the parameter b and number of recursions p;

– calculation of the H–distance d_p, $p = 0,1,2,\ldots,$ between the Heaviside function $h(t)$ and the activation functions $f_0, f_2, f_2, \ldots, f_p$ and activation functions $g_0, g_1, g_2, \ldots, g_p$;

– software tools for animation and visualization.

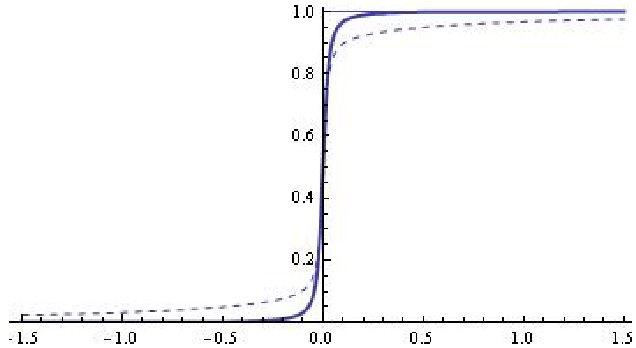

Fig. 6. Comparison of the functions $f_2(t)$ (dashed) and $g_2(t)$ (thick) for $b = 0.15$.

Some comparison of the functions $f_2(t)$ and $g_2(t)$ for $b = 0.15$ is plotted on Fig. 6.

From Fig. 6 we can conclude that each element of the sequence g_i gives a better "saturation" compared with the corresponding elements of the sequence f_i.

The Hausdorff approximation of the interval step function by the logistic and other sigmoidal functions is discussed from various approximation, computational and modelling aspects in [28]–[50].

References

[1] R. Alt, S. Markov, Theoretical and Computational Studies of some Bioreactor Models, Computers and Mathematics with Applications, 64, 2012, 350–360.

[2] G. Lente, Deterministic Kinetics in Chemistry and Systems Biology, Springer, New York, 2015.

[3] N. Kyurkchiev, S. Markov, On the numerical solution of the general kinetic K–angle reaction system, J. Math. Chem., 54 (3), 2016, 792–805.

[4] Z. Quan, Z. Zhang, The construction and approximation of the neural network with two weights, J. Appl. Math., 2014; doi:10.1155/2014/892653

[5] S. Adam, G. Magoulas, D. Karras, M. Vrahatis, Bounding the search space for global optimization of neural networks learning error: an interval analysis approach, J. of Machine Learning Research, 17, 2016, 1–40.

[6] N. Guliyev, V. Ismailov, A single hidden layer feedforward network with only one neuron in the hidden layer san approximate any univariate

function, Neural Computation, 28, 2016, 1289–1304.

[7] D. Costarelli, R. Spigler, Approximation results for neural network operators activated by sigmoidal functions, Neural Networks, 44, 2013, 101–106.

[8] D. Costarelli, G. Vinti, Pointwise and uniform approximation by multivariate neural network operators of the max–product type, Neural Networks, 2016, doi:10.1016/j.neunet.2016.06.002

[9] D. Costarelli, R. Spigler, Solving numerically nonlinear systems of balance laws by multivariate sigmoidal functions approximation, Computational and Applied Mathematics 2016, doi:10.1007/s40314–016–0334–8

[10] D. Costarelli, G. Vinti, Convergence for a family of neural network operators in Orlicz spaces, Mathematische Nachrichten 2016; doi:10.1002/mana.20160006

[11] J. Dombi, Z. Gera, The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators, Fuzzy Sets and Systems, 154 (2), 2005, 275–286.

[12] I. A. Basheer, M. Hajmeer, Artificial Neural Networks: Fundamentals, Computing, Design, and Application, Journal of Microbiological Methods, 43, 2000, 3–31, doi:10.1016/S0167–7012(00)00201–3

[13] Z. Chen, F. Cao, The Approximation Operators with Sigmoidal Functions, Computers & Mathematics with Applications 58, 2009, 758–765, doi:10.1016/j.camwa.2009.05.001

[14] Z. Chen, F. Cao, The Construction and Approximation of a Class of Neural Networks Operators with Ramp Functions, Journal of Computational Analysis and Applications, 14, 2012, 101–112.

[15] Z. Chen, F. Cao, J. Hu, Approximation by Network Operators with Logistic Activation Functions, Applied Mathematics and Computation, 256, 2015, 565–571, doi:10.1016/j.amc.2015.01.049

[16] N. Kyurkchiev, Mathematical Concepts in Insurance and Reinsurance. Some Moduli in Programming Environment MATHEMATICA, LAP LAMBERT Academic Publishing, Saarbrucken, 2016, 136 pp.

[17] N. Kyurkchiev, A. Andreev, Approximation and antenna and filter synthesis: Some moduli in programming environment Mathematica, LAP LAMBERT Academic Publishing, Saarbrucken, 2014, ISBN 978–3–659–53322–8.

[18] N. Kyurkchiev, S. Markov, Sigmoid functions: Some Approximation and Modelling Aspects, LAP LAMBERT Academic Publishing, Saarbrucken, 2015, ISBN 978–3–659–76045–7.

[19] D. Costarelli, R. Spigler, Constructive Approximation by Superposition of Sigmoidal Functions, Anal. Theory Appl., 29, 2013, 169–196, doi:10.4208/ata.2013.v29.n2.8

[20] D. Elliott, A better activation function for artificial neural networks, the National Science Foundation, Institute for Systems Research, Washington, DC, ISR Technical Rep. TR–6, 1993, http://ufnalski.edu.pl/zne/ci2014/papers/ElliottTR93–8.pdf

[21] K. Babu, D. Edla, New algebraic activation function for multi–layered feed forward neural networks. IETE Journal of Research, 2016, doi:10.1080/03772063.2016.1240633

[22] N. Johnson, S. Kotz, N. Balakrishnan (1994) Continuous Univariate Distributions, Vol. 1, NY, Willey.

[23] M. McLaughlin, Compendium of Common Probability Distributions, 2013, www.causascientia.org/Compendium.pdf

[24] F. Hausdorff, Set Theory (2 ed.) (Chelsea Publ., New York, (1962 [1957]), Republished by AMS–Chelsea, 2005, ISBN: 978–0–821–83835–8.

[25] B. Sendov, Hausdorff Approximations, Kluwer, Boston, 1990, doi:10.1007/978–94–009–0673–0

[26] Kyurkchiev, N., Initial approximation and root finding methods, WILEY–VCH Verlag Berlin GmbH, 104, 1998, 180 pp., ISBN 3–527–40132–6.

[27] N. Pavlov, N. Valchanov, N. Kyurkchiev, Bounds for the unique positive root of a polynomial in mathematics of finance, Plovdiv Univ., Sci. Works – Math., 37(3), 2010, 101–110.

[28] N. Kyurkchiev, A family of recurrence generated sigmoidal functions based on the Verhulst logistic function. Some approximation and modeling aspects, Biomath Communications, 3 (2), 2016, 18 pp.

[29] N. Kyurkchiev, S. Markov, Hausdorff Approximation of the Sign Function by a Class of Parametric Activation Functions, Biomath

Communications, 3(2), 2016, 14 pp., doi:10.11145/bmc.2016.12.217

[30] N. Kyurkchiev, On the Approximation of the step function by some cumulative distribution functions, Compt. rend. Acad. bulg. Sci., 68(12), 2015, 1475–1482.

[31] N. Kyurkchiev, S. Markov, On the Hausdorff distance between the Heaviside step function and Verhulst logistic function. J. Math. Chem., 54(1), 2016, 109–119, doi:10.1007/S10910–015–0552–0

[32] N. Kyurkchiev, S. Markov, Sigmoidal functions: some computational and modelling aspects. Biomath Communications, 1(2), 2014, 30–48, doi:10.11145/j.bmc.2015.03.081

[33] N. Kyurkchiev, S. Markov, On the approximation of the generalized cut function of degree $p+1$ by smooth sigmoid functions. Serdica J. Computing, 9(1), 2015, 101–112.

[34] N. Kyurkchiev, A note on the new geometric representation for the parameters in the fibril elongation process, Compt. rend. Acad. bulg. Sci., 69(8), 2016, 963–972.

[35] N. Kyurkchiev, S. Markov, A. Iliev, A note on the Schnute growth model, Int. J. of Engineering Research and Development, 12(6), 2016, 47–54.

[36] V. Kyurkchiev, N. Kyurkchiev, On the Approximation of the Step function by Raised–Cosine and Laplace Cumulative Distribution Functions, European International Journal of Science and Technology, 4(9), 2016, 75–84.

[37] A. Iliev, N. Kyurkchiev, S. Markov, On the Approximation of the Cut and Step Functions by Logistic and Gompertz Functions, BIOMATH, 4(2), 2015, 1510101, doi:10.11145/j.biomath.2015.10.101

[38] A. Iliev, N. Kyurkchiev, S. Markov, On the Approximation of the step function by some sigmoid functions, Mathematics and Computers in Simulation, 2015, doi:10.1016/j.matcom.2015.11.005

[39] N. Kyurkchiev, A. Iliev, A note on some growth curves arising from Box–Cox transformation, Int. J. of Engineering Works, 3(6), 2016, 47–51.

[40] N. Kyurkchiev, A. Iliev, On some growth curve modeling: approximation theory and applications, Int. J. of Trends in Research and Development, 3(3), 2016, 466–471.

[41] A. Iliev, N. Kyurkchiev, S. Markov, Approximation of the cut function by Stannard and Richards sigmoid functions, IJPAM, 109(1), 119–128, 2016.

[42] N. Kyurkchiev, A. Iliev, On the Hausdorff distance between the shifted Heaviside function and some generic growth functions, Int. J. of Engineering Works, 3(10), 2016, 73–77.

[43] N. Kyurkchiev, S. Markov, Approximation of the cut function by some generic logistic functions and applications, Advances in Applied Sciences, 1(2), 2016, 24–29.

[44] A. Iliev, N. Kyurkchiev, S. Markov, On the Hausdorff Distance Between the Shifted Heaviside Step Function and the Transmuted Stannard Growth Function, BIOMATH, 5(2), 2016, 1–6.

[45] N. Kyurkchiev, A. Iliev, On the Hausdorff distance between the Heaviside function and some transmuted activation functions, Mathematical Modelling and Applications, 2(1), 2016, 1–5.

[46] N. Kyurkchiev, Uniform Approximation of the Generalized Cut Function by Erlang Cumulative Distribution Function. Application in Applied Insurance Mathematics, International Journal of Theoretical and Applied Mathematics, 2(2), 2016, 40–44.

[47] N. Kyurkchiev, S. Markov, Hausdorff approximation of the sign function by a class of parametric activation functions, Biomath Communications, 3(2), 2016, 14 pp.

[48] V. Kyurkchiev, N. Kyurkchiev, A family of recurrence generated functions based on the "half–hyperbolic tangent activation function", Biomedical Statistics and Informatics, 2017 (accepted).

[49] N. Kyurkchiev, A. Iliev, S. Markov, Families of Recurrence Generated Three and Four Parametric Activation Functions, Int. J. Sci. Res. and Development, 4(12), 2017, 746–750.

[50] A. Iliev, N. Kyurkchiev, S. Markov, A family of recurrence generated parametric activation functions with applications to neural networks, International Journal on Research Innovations in Engineering Science and Technology, 2(1), 2017, 60–68.