ISSN: 0067-2904

Monotone Approximation by Quadratic Neural Network of Functions in L^p Spaces for p<1

Hawraa Abbas Almurieb*, Eman Samir Bhaya
Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Hillah, Iraq

Received: 30/7/ 2019 Accepted: 30/9/2019

Abstract

Some researchers are interested in using the flexible and applicable properties of quadratic functions as activation functions for FNNs. We study the essential approximation rate of any Lebesgue-integrable monotone function by a neural network of quadratic activation functions. The simultaneous degree of essential approximation is also studied. Both estimates are proved to be within the second order of modulus of smoothness.

Keywords: Essential approximation, monotone approximation, modulus of smoothness, activation function, neural network.

التقريب الأساسي الحافظ للرتابة باستخدام الشبكات العصبية ذات دالة تفعيل تربيعية في الفضاء L^p

حوراء عباس المرعب*, إيمان سمير بهية
قسم الرياضيات ، كلية التربية للعلوم الصرفة ، جامعة بابل ، الحلة ، العراق

الخلاصة

نظراً لما تتمتع به الدوال التربيعية من خصائص تطبيقية مرنة، فقد قام الباحثون باستخدامها في العديد من الشبكات العصبية التقدمية. قمنا في هذا البحث بدراسة درجة التقريب الأساسي لدوال رتيبة قابلة للتكامل ليبيكياً بواسطة شبكات عصبية ذات دالة تفعيل تربيعية. كما قمنا بدراسة درجة التقريب المتزامن لتلك الشبكات، لنحصل بذلك على درجة تقريب متعلقة بالمقياس الثاني للنعومة لكلا الحدين.

1. Introduction

In the field of Artificial Neural Networks, researchers can use the neural network to approximate a function with an acceptable close degree to the original function.

In 1989, Cybenko [1] introduced the Universal Theorem of Approximation, which states that for any continuous function f defined on a compact space X, there exists a neural network $N_n f$ such that

$$\|f - N_n f\| < \epsilon$$

Other researchers concluded similar results in that period [2, 3].

In spite of the importance of the above theorem, it suffered from a number of limitations, including being primitive for the following aspects; the used function space, the degree of approximation and the nature of the neural network itself. Cybenko took the function f from a very tight space. Other spaces were studied later such as Banach Space [4], Sobolov space [5], Hilbert spaces [6], and other wider spaces [7, 8].

*Email:pure.hawraa.abbas@uababylon.edu.iq

In addition, the activation function of the neural network defined above by Cybenko is not specified. He proved it for any neural network with any arbitrary activation function. The importance of the activation function is not only to create a relationship between inputs and outputs but also to add the ability to the network to learn any type of data. To build a more powerful network, it is essential to choose a suitable activation function depending on various issues, such as type of data, number of hidden layers, and the network's model. Sigmodal, threshold, binary, identity, tan and arctan are examples of some types of functions that activate the neural network.

Cybenko's Theorem was for any arbitrary activation function. Later, different types of neural networks with specified activation functions were defined to achieve good approximation [9-20].
For deep learning applications, the squared activation functions were used efficiently in different areas to achieve favourable properties, which generated expressive networks with good learning abilities [21]. Here, we define the quadratic activation function as

$$\varphi(x) = x^2 , x \in [0,1]. \quad (1)$$

Divide [0,1] into n subintervals I_k each of length at most $\delta = \frac{-1-2(k-1)}{n}, \forall\ k = 1, \dots, n$. So for any $i \neq k, i, k = 1, 2, \dots, n$, we define

$$S_n(x_k) = \frac{c_n}{n} \sum_{i=1}^{n} [f(x_k + x_i) + f(x_k - x_i) - 3f(x_k)]\varphi^i(x_k) \quad (2)$$

which satisfies that

$$c_n \sum_{i=1}^{n} S_n(x_i) = 1 \quad (3)$$

If $f \in L_p[0,1], p < 1,$ and for some $d < n$, the following neural network operator is defined

$$N_{n,d}(f, x_k) = \sum_{j=0}^{d} f\left(\frac{j}{n}\right) S_n(x_k), \text{for } x_k \in I_k \quad (4)$$

We name Ω to be the set of all neural networks of type **Error! Reference source not found.**.
Now, we move to the criterion of approximation that makes the rate of approximation as accurate as possible.

$$E_n(f)_p = \inf_{N \in \Omega} \|f - N\|_p \quad (5)$$

The rate of monotone approximation of a function f on $L_p(I)$ by elements of Ω is denoted by E(I). Modulus of smoothness is the best measure of the rate of approximation, since it judges the accuracy of the best approximation of a function as the error of approximation is estimated. Smoothness of functions can be measured by the following modulus [22]:

$$\omega_k(f, \delta)_p = \sup_{0<h\leq\delta} \|\Delta_h^r f(\cdot)\|_p, \delta \geq 0, \quad (6)$$

which is called the rth modulus of smoothness, where the rth symmetric difference of f is

$$\Delta_h^r(f, x) := \begin{cases} \sum_{i=0}^{r} \binom{r}{i} (-1)^{r-i} f\left(x - \frac{rh}{2} + ih\right), & x \pm \frac{rh}{2} \in I \\ 0, & o.w \end{cases} \quad (7)$$

2. Auxiliary Lemmas

At the beginning, we need some useful facts about the relationships among moduli of smoothness and/or the rate of approximation. These facts were previously proved in details by other authors [23-25]

Lemma 1

The first inequality gives a relationship between the error of approximation, while the second is a more general one

$$E(I) \leq c\omega_3(f, |I|, I)_p \quad (8)$$

$$E_{r-1}(I) \leq c\omega_r(f, |I|, I)_p \quad (9)$$

Lemma 2

To derive S_n in (**Error! Reference source not found.**) , let $m = 1,2, \dots, r$, then

$$S_n^{(m)} = \frac{c_n}{n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \binom{r}{i} [f^{(m)}(x_k + x_i) + f^{(m)}(x_k - x_i) - f^{(m)}(x_k)]\emptyset^i(x_k) + \frac{c_n}{n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \binom{r}{i} [f(x_k + x_i) + f(x_k - x_i) - f(x_k)]\emptyset^{i^{(m)}}(x_k) \quad (10)$$

which easily implies that the derivative of the operator **Error! Reference source not found.** is:

$$N_{n,d}^{(m)}(f,x_k)=\sum_{j=1}^{d}c_j f\left(\frac{j}{n}\right)S_n^{(m)}(x_k) \tag{11}$$

3. Main Results

Let f be an increasing function in L_p, and since $\varphi^i, \forall\ i=1,\dots,n$, is increasing on [0,1] as well, then so is $N_{n,d}(f,\cdot)$, for any $N_{n,d}\in\Omega$. So we conclude a monotone approximation for the function f to be equivalent to the second modulus of smoothness for both upper and lower bounds.

We begin with the following main result for the upper bound of $E_n(f)$.

Theorem 1

If $f\in L_p[0,1], p<1$, then there exists $N_{n,d}\in\Omega$, such that

$$\left\|f-N_{n,d}(f)\right\|_p^p\le\frac{c_n c(p)}{n}\left[\omega_2(f,I_k)_p^p+\|f\|_p^p\right].$$

Proof

By (**Error! Reference source not found.**) and (**Error! Reference source not found.**),

$$\left\|f-N_{n,d}\right\|_p^p=\sum_{i=0}^{n}\left|f(x_i)-N_{n,d}(x_i)\right|^p$$

$$\le\sum_{\substack{i,k=1\\ i\neq k}}^{n}\sum_{j=1}^{d}\frac{c_j c_n}{n}\left|f(x_k+x_i)+f(x_k-x_i)-2f(x_k)+f\left(\frac{j}{n}\right)\right|^p x_k^{2ip}$$

$$\le\sum_{\substack{i,k=1\\ i\neq k}}^{n}\sum_{j=1}^{d}\frac{c_j c_n}{n}\left|f(x_k+x_i)+f(x_k-x_i)-2f(x_k)\right|^p\sum_{i=0}^{\infty}x_k^{2ip}+\sum_{j=1}^{d}\left|f\left(\frac{j}{n}\right)\right|^p\sum_{i=0}^{\infty}x_k^{2ip}$$

By using Taylor Series of $\sum_{i=0}^{\infty}\left(x_k^2\right)^i$ for $|x_k|<1$, we have

$$\left\|f-N_{n,d}\right\|_p^p\le\sum_{j=1}^{d}\frac{c_j c_n}{n}\left(\sum_{i=0}^{\infty}\left(x_k^2\right)^i\right)^p\left[\Delta_h^2(f,I_k)+\|f\|_p^p\right]$$

$$\le\frac{c_n}{n}\left(\frac{1}{1-x_k^2}\right)^p\left[\omega_2(f,|I_k|)_p^p+\|f\|_p^p\right]$$

$$\le\frac{c_n c(p)}{n}\left[\omega_2(f,I_k)_p^p+\|f\|_p^p\right].\blacksquare$$

Also, we can prove the lower bound to get an essential approximation to the function f by $N_{n,d}$.

Theorem 2

If $f\in L_p[0,1], p<1$, then there exists $N_{n,d}\in\Omega$, with $d<n$, such that

$$\omega_2(f,|I_k|)_p^p\le\frac{c(p)}{n^{p+1}}\left[\|f\|_p^p+E_n(f)_p^p\right]$$

Proof

Let $b=max\{i;2^i<n\}$, then we get the expansion of $N_{n,d}$ in **Error! Reference source not found.** as follow

$$N_{n,d}(x)-N_{0,d}(x)$$
$$=N_{n,d}(x)-N_{2^b,d}(x)+\cdots+\left(N_{2^b,d}(x)-N_{2^{b-1},d}(x)\right)+\cdots+\left(N_{1,d}(x)-N_{0,d}(x)\right)$$

For $m<n$, suppose that

$$\|N_n-N_m\|_p^p\le 2c(p)E_m(f)_p^p,$$

Then,

$$\omega_2(f,\delta)_p^p\le c(p)\left[\omega_2(f-N_n,\delta)_p^p+\omega_2(N_n,\delta)_p^p\right]$$
$$\le c(p)\left[\|f-N_n\|_p^p+\omega_2(N_n,\delta)_p^p\right]$$

$$\leq \frac{c(p)}{n^p}\left[n^{-1}\left(\omega_2\left(f,\delta\right)_p^p + \|f\|_p^p\right) + \omega_2\left(N_{2^i} - N_{2^{i-1}}, 2^{-b}\right)_p^p\right]$$

$$\leq \frac{c(p)}{n^{p+1}}\left[\omega_2\left(f,\delta\right)_p^p + \|f\|_p^p + \sum_{m=2}^{n}(m+1)^{p-1}E_m(f)_p^p\right]$$

$$\leq \frac{c(p)}{n^{p+1}}\left[\|f\|_p^p + E_n(f)_p^p\right]$$

So, we get

$$\omega_2\left(f,\delta\right)_p^p \leq \frac{c(p)}{n^{p+1}}\left[\|f\|_p^p + E_n(f)_p^p\right]. \blacksquare$$

Now, for derivatives of the function and its operator, we find that

Theorem 3

If $f \in L_p^r[0,1], p < 1$, then for any $m = 0,1,\ldots r$, there exists $N_{n,d} \in \Omega$, with $d < n$, that satisfies the following

$$\left\|f^{(m)} - N_{n,d}^{(m)}(f)\right\|_p^p \leq \frac{c}{n^d}\left[\left(1 + \frac{1}{n^m}\right)\omega_2\left(f^{(m)}, |I_k|\right)_p^p + 2\|f\|_p^p\right]$$

Proof

Using the derivatives of (**Error! Reference source not found.**) and **Error! Reference source not found.** located in (**Error! Reference source not found.**) and (**Error! Reference source not found.**), respectively, we have

$$\left\| N_{n,d}^{(m)} - f^{(m)}\right\|_p^p$$

$$\leq \sum_{j=1}^{d}\frac{c_j c_n}{n}\left[\sum_{\substack{i,k=1\\ i\neq k}}^{n}\binom{m}{i}\left[f^{(m)}\left(x_k + x_i\right) + f^{(m)}\left(x_k - x_i\right) - 2f^{(m)}\left(x_k\right) + f\left(\frac{j}{n}\right)\right]\varphi^{ip}\left(x_k\right)\right.$$

$$+ \sum_{\substack{i,k=1\\ i\neq k}}^{n}\binom{m}{i}\left[f\left(x_k + x_i\right) + f\left(x_k - x_i\right) - f\left(x_k\right) - f^{(m)}\left(x_k\right)\right.$$

$$\left.\left. + f\left(\frac{j}{n}\right)\right]\varphi^{ip^{(m)}}\left(x_k\right)\right]^p$$

Using the same techniques as the last proof, we get

$$\left\|N_{n,d}^{(m)} - f^{(m)}\right\|_p^p \leq \sum_{j=1}^{d}\frac{c_j c_n c(p)}{n}\binom{m}{i}\left[\left(1 + \frac{1}{n^m}\right)\omega_2\left(f, |I_k|\right)_p^p + 2\|f\|_p^p\right]$$

$$\leq \frac{C}{n^d}\left[\left(1 + \frac{1}{n^m}\right)\omega_2\left(f, |I_k|\right)_p^p + 2\|f\|_p^p\right]. \blacksquare$$

It is easy to conclude that the essential approximation also holds for derivatives of any $f \in L_p^r[0,1], p < 1$.

Corollary 4

If $f \in L_p^r[0,1], p < 1$, then for any $m = 0,1,\ldots r$, there exists $N_{n,d} \in \Omega$, with $d < n$, that satisfies the following

$$\omega_2\left(f^{(m)}, |I_k|\right)_p^p \leq \frac{c(p)}{n^{p+1}}\left[\left\|f^{(m)}\right\|_p^p + E_n\left(f^{(m)}\right)_p^p\right].$$

References

1. Cybenko, G. **1989.** Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems*, **2**(4): 303-314.
2. Funahashi, K.I. **1989.** On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural networks*, **2**(3): 183-192.

3. Hornik, K. **1993.** Some New Results on Neural Network Approximation. *Neural networks*, **6**(8): 1069-1072.

4. Chen, T. and Chen, H. **1995.** Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and its Application to Dynamical Systems, IEEE *Transactions on Neural Networks*, pp: 911-917.

5. Petrushev, P. P. **1998.** Approximation by Ridge Functions and Neural Networks. *SIAM Journal on Mathematical Analysis*, **30**(1): 155-189.

6. Jones et al, L. K. **1992**. A Simple Lemma on Greedy Approximation in Hilbert Space and Convergence Rates for Projection Pursuit Regression and Neural Network Training. *The annals of Statistics*, **20**(1): 608-613.

7. Costarelli, D. **2015**. Neural network Operators: Constructive Interpolation of Multivariate functions. *Neural Networks*, **67**: 28-36.

8. Donahue, M. J. Darken, C., Gurvits, L. and Sontag, E. **1997.** Rates of Convex Approximation in Non-hilbert Spaces. *Constructive Approximation*, **13**(2): 187-220.

9. Shin, Y. and Ghosh, J. **1991.** The pi-sigma Network: An Efficient Higher-order Neural Network for Pattern Classification and Function Approximation. In IJCNN-91-Seattle International Joint Conference on Neural Networks, IEEE, volume 1, pp: 13-18.

10. Barron, A. **1993.** Universial Approximation Bounds for Superpositions of a Sigmodial Function, IEEE Trans., *Inform. Theory*, **39**(3): 930-945.

11. Suzuki, S. **1998**. Constructive Function Approximation by Three-Layer Artificial Neural Networks. *Neural Networks*, **11**(6): 1049-1058.

12. Cao, F., Xu, Z. and Li, Y. **2005.** Pointwise Approximation for Neural Networks. *In International Symposium on Neural Networks*, Springer, pp: 39-44.

13. Xu, Z. B. and Cao, F. L. **2005**. Simultaneous lp-approximation Order for Neural Networks. *Neural Networks*, **18**(7): 914-923.

14. Ding, C. Cao, F. and Xu, Z. **2006**. *The Essential Approximation Order for Neural Networks with Trigonometric Hidden Layer Units*, Springer-Verlag Berlin Heidelberg, **3971**: 72-79.

15. Chen, Z. and Cao, F. **2009**. The Approximation Operators with Sigmoidal Functions, *Comput. Math. Appl*., **58**(4): 758-765.

16. Costarelli, D. and Vinti, G. 2016. Pointwise and Uniform Approximation by Multivariate Neural Network Operators of the Max-product Type. *Neural Networks*, **81**: 81-90.

17. Almurieb H. **2017**. Simultaneous Approximation of Order m by Artificial Neural Networks, *International Journal of Applied Mathematics and Statistics*, **56**(4): 38-44.

18. Asdrubali, F., Baldinelli, G., Bianchi, F.,Costarelli, D., Rotili, A., Seracini, M. and Vinti, G. **2018.** Detection of thermal bridges from thermographic images by means of image processing approximation algorithms. *Applied Mathematics and Computation*, **317**: 160-171.

19. Petersen, P. and Voigtlaender, F. **2018**. Optimal Approximation of Piecewise Smooth Functions Using Deep Relu Neural Networks. *Neural Networks*, **108**: 296-330.

20. Bhaya, E.S. and Fadel, Z.M. **2018.** Nearly Exponential Approximation for Neural Networks. *Journal of University of Babylon*, **26**(1): 103-113.

21. Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S. **2018.** Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. arXiv preprint arXiv:1811.03378.

22. Ditzian, T. and Totik, V. **1987.** *Moduli of Smoothness*, Springer-Verlag, Berlin Heidelberg, New York.

23. Bhaya, E.S. **2003**. On the Constrained and Unconstrained Approximation. PhD thesis, University of Baghdad, Baghdad, Iraq.

24. DeVore, R.A. **1976**. Degree of Approximation. *Approximation theory II*, **241**(242): 117-161.

25. DeVore, R. A. and Lorentz, G.G. **1993.** *Constructive Approximation*, volume 303. Springer Science and Business Media.