MJPAS Vol. 1, No. 2 (2023) 37-45

Mustansiriyah Journal of Pure and Applied Sciences

www.MJPAS.com

L_P Approximation by Fixed Weighted Neural Network

S. S. Mahdi [1,*] , E. S. Bhaya [1, 2]

[1,2] Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Iraq. (sara.salih638@gmail.com)
[2]Department of Mathematics, College of Education, Al-Zahraa University for Women, Iraq. (emanbhaya@alzahraa.edu.iq, emanbhaya@uobabylon.edu.iq.)

A B S T R A C T

Neural networks are tools for parallel computations of functions with several variables, it is a linear computation of non-linear functions that we are call activation functions. In our paper we deal with one hidden layer feed forward neural network.

We have two problems with neural networks approximation, the first is the density problem concerns. With the condition for the approximated function to approximate by the neural networks. In previous works on approximation using neural networks, the weights are different for each input. This makes engineering applications very difficult.

In this work we approximate any bounded function on L_p for $p < 1$ by a forward neural network and find the degree of the best approximation by this neural network using the k' th order of smoothness. Then we connect neurons number and the degree of the best approximation.

Keywords: L_p *norm, Neural network,* Approximated, Approximated function, *Best approximation***.**

1. Introduction

Firstly, let us introduce some basics notations and definitions that we need in our work. Begin with the best approximation to F from Υ is F^* satisfy

$$\|f - f^*\| < inf_{y\in\Upsilon}\|f - \Upsilon\|$$

Define

$$L_p(R) = \{f: R \to R: \|f\|_p < \infty\}$$

Where:

$$\|f\|_p = (\int_R f\,|(x)|^p dx)^{\frac{1}{p}} \qquad 1 \le p < \infty.$$

The $k'th$ order ordinary modulus of smoothness define as

$$w_k(f,\delta)_p = sup_{|h|\le\delta}\|\Delta_h^k(f,\delta)\|_p \quad ,\delta \ge 0.$$

$$w_k(f,\delta)_\infty = sup_{|h|\le\delta}\left|\Delta_h^k(f,\delta)\right| \quad for \quad 0 < p \le \infty.$$

Where:

$$\Delta_h^k(f,\delta) = \sum_{i=0}^{k}(\, i^k\,) f\,(x+ih)(-1)^{k-i}.$$

The degree of the best approximation of $f \in L_p(R)$ from the subspace G is

$$E_n(f)_p = inf_{g\in G}\|f-g\|_p.$$

Neural networks are tools for parallel computations of functions with several variables. It is a linear computation of non-linear functions that we are call activation functions, in our paper, we dealt with neural networks, which have 3-layers input layer, hidden layer and output layer.

In general, we can define the neural network mathematically as

$$H(x) = \sum_{i=1}^{N} c_i\sigma(K.x+\varphi_i),$$

Where, σ is the activation function with one variable

x, $k \in R^s$ are weights and c_i are thresholds and $\varphi_i \in \mathbb{R}$.

Let us now recall some examples [1] of activation functions

$\sigma(x) = (1+e^{-x})^{-1}$ [the squashing function].

$\sigma(x) = (1+x^2)^{\alpha}$ [generalized multiquadrics] $where\ \alpha \notin Z.$

$\sigma(x) = |x|^{2q-1}$ [thin plate splines] $q \in N$.

We can define the sigmodal function as

$$\lim_{x\to\infty} \sigma(x) = 1$$

$$\lim_{x\to-\infty} \sigma(x) = 0 \ .$$

We have two problems with neural networks approximations, first problem is called density problem, deals with the approximated function to approximate by the neural networks.
The complexity is analogue with the problem of the best approximation.
In [2-9], the authors studied the approximation of a continuous function defined on compact set in $\mathbb{R}$, using neural network with sigmodal activation function.

In [6] and [5] Funhashi and Cybenko find if f is continuous function it is domain is compact set in $\mathbb{R}^s$ then we can approximate f by a forward neural network. This neural network has an infinite number of neurons.

In [4] Chui and Li demonstrated that the sigmodal activation functions can contain integer weights and that the threshold can be an approximation to any continuous function defined on a compact subset of $\mathbb{R}$.

In [3] Chen and Chen improved the results of Funhashi and Cybenko who showed that any continuous function on compact subset in $\mathbb{R}$s can be approximated by a feedforward neural network with a bounded sigmodal function and need not be continuous.

In the articles [3-6] weights are variant in the neural network approximation (each x has a weight), for different inputs, this makes engineering applications very difficult.

In this work we approximate any bounded function on L_p for $p < 1$ by a forward neural network and find that the degree of the best approximation by this neural network using the $k'th$ order of smoothness. Then we find a relation between the number of neurons and the degree of the best approximation.

2. The Main Results

In this section, we shall define a neural network polynomial has fixed weights, then we prove it is the best approximation of functions in $L_p(\mathbb{R})$, $0 < p < 1$, in terms of the first and the $K'th$ usual models of smoothness.

Theorem 2.1

σ is a bounded function of sigmoidal type, defined on $\mathbb{R}$, if $f \in L_p(\mathbb{R})$ with $\to 0$ as $x \to \infty$, then there exists constants $\varphi_i, c_i \in \mathbb{R}$ and positive integer K, N such that

$$\|f - H\|_p < C(p)\, w_k(f, \delta)_p,$$

Where:

$$H(x) = \sum_{i=1}^{N} c_i \sigma(K.x + \varphi_i).$$

Proof

Let us divide $[-N, N]$ as 2 N^2, since σ a sigmoidal function then there exists r $\in \mathbb{R}$

$$\|\sigma(x)\|_p < \frac{1}{N^2} \quad for\ x \geq r$$

And

$$\|\sigma(x)\|_p < \frac{1}{N^2} \ for\ x \leq -r\,.$$

Choose $K\ E\ Z^{+,}$ so there exist, $\frac{K}{(2N)} > r$.

Now let us write our neural network

$$B(x) = \sum_{i=1}^{2N^2} \sum_{j=1}^{K} (j^k)(-1)^{K-j} f(x + jh)\sigma(K(x - \varphi_i)).$$

If $x \leq -N \quad \Rightarrow K(x - \varphi_i) \leq -r$ and hence $\|f\|_p \to 0 \quad as\ x \to \infty$ we can find $l \in N$ as

$$\|f(x)\|_p < w_k(f, \delta)_p \leq c w_k(f, \delta)_p \quad \forall x \in l. \tag{1}$$

Where c is the absolute constant, then

$$\|B(x)\|_p = \| \sum_{i=1}^{2N^2} \sum_{j=1}^{K} (j^k)(-1)^{K-j} f(x_0 + jh)\sigma(K(x - \varphi_i)\|_p \tag{2}$$

Since

$$\lim_{x\to\infty} \sigma(x) = 1$$

$$\lim_{x\to-\infty} \sigma(x) = 0.$$

So σ is bounded and

$$\|B(x)\|_p \leq c(p) \sum_{i=1}^{2N^2} \left(\int_R \left| \sum_{j=1}^{K} (j^k)(-1)^{K-j} f(x_0 + jh) \right|^p |\sigma(K(x - \varphi_i)|^p\, dx\right)^{1/p}$$

$$\leq c(p) N^2 \frac{1}{N^2} w_k(f, \delta)_p.$$

So

$$\|f(x) - B(x)\|_p \leq c(p) \left(\|f\|_p + \|B(x)\|_p\right) \leq c(p) w_k(f, \delta)_p.$$

If $x \in [-N, N]$, then $x \in [x_{i_0} - 1, x_{i_0}]$ for some i_0, with $1 < i_0 < 2N^2$. Note that

$$K(x - \varphi_i) \geq r \quad for\ \ i = 1, \dots, i_0 - 1$$

And

$$K(x - \varphi_i) \leq -r \quad for\ \ i = i_0 + 1, \dots, 2N^2.$$

Then from the fact

$$\sum_{i=1}^{i_0-1} (j^k)f(x+ih)(-1)^{k-i}\sigma(k(x-\varphi_i)$$

$$=\sum_{i=1}^{i_0-1} (j^k)f(x+ih)(-1)^{k-i}\sigma(k(x-\varphi_i)-1)-f(x_{i0}-1)+f(x_0).$$

We have

$$\|f(x)-B(x)\|_p=\|f(x)-\sum_{i=1}^{2N^2}\sum_{j=1}^{K} (j^k)(-1)^{K-j}f(x_0+jh)\sigma(K\big(x-\varphi_j)\big)\|_p.$$

Using (1) and (2), we get

$$\|f-B\|_p\le c(p)\|f\|_p+\|B\|_p\le c(p)w_k(f,\delta)_p.$$

If $x\ge N$, then $k(x-\varphi_i)\ge r$ and hence

$$\|\sigma\big(k(x-\varphi_i)\big)-1\|_p\le\frac{1}{N^2},\qquad for\ i=1,\dots,2N^2.$$

Then

$$\|f(x)-B(x)\|_p=\|f(x)-\sum_{i=1}^{2N^2}\sum_{j=1}^{K} (j^k)(-1)^{K-j}f(x+jh)\sigma\big(K(x-\varphi_i)\big)-1$$
$$+\sum_{i=1}^{2N^2}\sum_{j=1}^{K} (j^k)(-1)^{K-j}f(x+jh)\|_p$$

$$\le c(p)\|f(x)\|_p+\|f(N)\|_p+\|f(-N)\|_p+\sum_{j=1}^{k} \|\big(j^k\big)f(x+jh)(-1)^{k-j}\sigma\big(k(x-\varphi_i)\big)-1\|_p$$

$$\le c\ (p)\big(\|f\|_p+\|B\|_p\big).$$

Then using (1) and (2) to get

$$\|f-B\|_p\le c(p)w_k(f,\delta)_p$$

This ends the proof ■.

The convolution of two functions f and g defined on ℝ, can be defined as

$$(f*g)\ (x):=\int_R f(y)g\ (x-y)\ dy.$$

For $x \in \mathbb{R}$, we define

$$\begin{cases} ca^{\frac{3}{\sqrt{1-x}}}, & \text{if } |x| < 1 \\ 0, & \text{if } |x| \geq 1 \end{cases} \tag{3}$$

Where $a > 0$ and c is chosen so that $\int_R^3 G(x)dx = 1$ then all the derivatives of G are in $L_p(\mathbb{R})$. For each positive integer k, we define

$$G_k(x) = kG(kx) \tag{4}$$

Then

$$\int_R G_k(x)dx = 1 \; and \; G_k \in L_p(R).$$

For any positive integer k.

$$G_k * f \to f$$

$$\|G_k * f - f\|_p \to 0$$

In the following lemma we shall proof $G_k * f$ uniformly convergence to f.

Lemma 2.2

If $f \in L_p(\mathbb{R})$ then

$$\|G_k * f - f\|_p < w_1(f,\delta)_p \;, \qquad \delta > 0$$

Proof

Let $f \in L_p$ $(\mathbb{R})$

$$\|G_k * f - f\|_p = (\int_R \left|\int_{-1}^{1} kG(ky)(f(x-y)dy - f(x)\right|^p dx)^{\frac{1}{p}}$$

$$= \left(\int_R kG(ky)\big(f(x-y) - f(x)\big)dy\right).$$

Let $ky = z, dz = k.dy, y = \frac{z}{k}$

$$\|G_k * f - f\|_p = (\int_R \left|\int_{-1}^{1} G(z) \right|^p \left|f\left(x - \frac{z}{k}\right) - f(x) \right|^p dz)^{\frac{1}{p}}$$

$$\leq c(p)w_1(f,\delta)_p \, .$$

We shall use the convolution to prove the sufficiency of the boundedness fixed weight sigmoidal function with the approximation by neural network for f $\in L_p$, $0 < p < 1$ on compact set ■.

Theorem 2.3

Let $f \in L_p(\mathbb{R})$, if σ a bounded and measurable function on $\mathbb{R}$, then we can find $\varphi_i, c_i \in \mathbb{R}$ and $K, N \in z$ satisfy

$$\|f - H\|_p < c(p)w_k(f,\delta)_p$$

Where:

$$H(x) = \sum_{i=1}^{N} c_i\sigma(Kx + \varphi_i), a \leq x \leq b.$$

Proof

Let G and G_k are the same as in (3) and (4), then let us defin $\bar{f} \in L_p(\mathbb{R})$ such that $\bar{f} = f$ on $[a,b]$ and $\bar{f} = 0$ outside of $[a-1, b+1]$.
By Lemma 2.2 we get

$$G_k * \bar{f} \to \bar{f} \quad on\ [a,b].$$

Since

$$\int_R G_k(x-y)\bar{f}(y)\,dy < \infty \,, for\ each\ positive\ integer\ K,$$

$G_k * \bar{f}$ is approximated by a Riemann sum.
For any K∈Z$^+$ we can find $P_k \in Z$ and real constants $y_i, c_i \quad for\ i = 1, \dots, P_k$ satisfies

$$\|\left(G_k * \bar{f}\right)\ (x) - \sum_{i=1}^{P_k} c_i G_k\ (x-y)\bar{f}(y_i)\|_p \leq c\ (p)\, P_k \left(\frac{1}{P_k}\right) w_1\, f\ (f,\delta)_p \tag{5}$$

Where $y_i \in \mathbb{R}\,,\ for\ i = 1 \dots\dots, P_k$

Since $G_k \in L_p(\mathbb{R})$, by Theorem (2.1), there exists constants $\alpha_{j,k}, \beta_{j,k} \in \mathbb{R}$ and positive integer h such that

$$\|G_k(x-y_i) - \sum_{j,k} \beta_{j,k}\, \sigma\,(h(x-y_i) + \alpha_{j,k}\,)\,\|_p \leq c(p)w_k(f,\delta)_p. \tag{6}$$

Now, using Lemma (2.2) we choose a positive integer K satisfies

$$\|f-\left(G_k * \bar{f}\right)(x)\ \|_p \leq c(p) w_1(f, \delta)_p \text{, for } x \in R \tag{7}$$

From (5) and (7) we have

$$\|f(x)-\sum_{i=1}^{P_k} c_i \bar{f}(y_i) \sum_{j,k} \beta_{j,k}\, \sigma(h(x-y_i)+\alpha_{j,k})\|_p$$

$$\leq \|f(x)-\left(G_k * \bar{f}\right)(x)\|_p+\|\left(G_k * \bar{f}\right)(x)-\sum_{i=1}^{P_k} c_i G_k(x-y_i) \bar{f}(y)\|_p$$

$$+\left\|\sum_{i=1}^{P_k} c_i G_k(x-y_i) \bar{f}(y)-\sum_{i=1}^{P_k} c_i \bar{f}(y_i) \sum_{j,k} \beta_{j,k} \sigma(h\,(x-y_i)+\alpha_{j,k}\right\|_p$$

$$\leq w_1(f, \delta)_p+w_1(f, \delta)_p+w_k(f, \delta)_p$$

$$\leq c(p)\left(w_k(f, \delta)_p+w_1(f, \delta)_p\right) \blacksquare$$

3. Conclusion

We prove the density result by neural networks with a fixed weight. We conclude numerically that any sigmoidal neural network with a fixed weight can approximate any bounded function on L_p for $p<1$ by forward neural network and find that the degree of the best approximation by this neural network using the $K'th$ order of smoothness. We connect neurons number and the degree of the best approximation. This lead to that our approximation in this neural network is strong.

4. Future Work

Define a multilayer type neural network, then we study its approximation for function in $L_P[a,b]^a$. In term of k-multivariate modulus of smoothness.

5. References

[1] Hahm, N., & Hong, B. I. (1999). Extension of localized approximation by neural networks. *Bulletin of the Australian Mathematical Society*, *59*(1), 121-131.

[2] Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, *39*(3), 930-945.

[3] Chen, T., Chen, H., & Liu, R. W. (1992). A constructive proof and an extension of Cybenko's approximation theorem. In *Computing Science and Statistics: Statistics of Many Parameters: Curves, Images, Spatial Models*, Springer New York, pp. 163-168.

[4] Chui, C. K., & Li, X. (1992). Approximation by ridge functions and neural networks with one hidden layer. *Journal of Approximation Theory*, *70*(2), 131-141.

[5] Funahashi, K. I. (1998). Multilayer neural networks and Bayes decision theory. *Neural Networks*, *11*(2), 209-213.

[6] Ito, Y. (1991). Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Networks*, *4*(3), 385-394.

[7] Stinchcombe, M. B. (1999). Neural network approximation of continuous functionals and continuous functions on compactifications. *Neural Networks*, *12*(3), 467-477.

[8] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359-366.

[9] Ranjita R., (2018), A study of neural network based function approximation, *International Journal of Engineering and Science Invention*, *7*(1), 32-34