The Essential Order of $(L_p, p < 1)$ Approximation Using Regular Neural Networks

Eman S. Bhaya **Omar A. Al-sammak**
University of Babylon, College of Education for Pure Sciences, Department of Mathematics
emanbhaya@itnet.uobabylon.edu.iq omar.alhelly@itnet.uobabylon.edu.iq

Abstract

This paper is concerning with essential degree of approximation using regular neural networks and how a multivariate function in $L_p(K)$ spaces for $p < 1$ can be approximated using a forward regular neural network. So, we can have the essential approximation ability of a multivariate function in $L_p(K)$ spaces for $p < 1$ using regular FFN.

Keywords. Neural network approximation, Modulus of smoothness, $L_p(K)$ Spaces, best approximation.

الخلاصة

درسنا في هذا البحث درجة التقريب الاساسي بأستخدام الشبكة العصبية المنتظمة ، وكيف يمكن تقريب الدوال المتعددة المتغيرات في فضاء $L_p(K)$ عندما $p < 1$ بأستخدام الشبكة العصبية الامامية المنتظمة ، وكذلك بامكاننا الحصول على مبرهنات مباشرة وعكسية ونظرية تكافؤ للتقريب المتعددة المتغيرات في فضاء $L_p(K)$ عندما $p < 1$ بأستخدام الشبكة العصبية الامامية المنتظمة .

الكلمات المفتاحية : تقريب الشبكة العصبية ،مقياس النعومة ، فضاءات $L_p(K)$ ، التقريب الافضل .

1. Introduction

Various papers on feasibility of approximation by forward neural networks have been made in past years (see [Cardaliaguet & Euvrard,1992; Chen, 1995; Chen, 1994; Chui, 1992; Cybenko, 1989; Gallant, ,1992; Hornik, 1989; Hornik, & Stinchombe, 1990; Leshno *et al.*, 1993; Mhaskr & Michelli,1992].

The most important result among these papers is that :

If we have a continuous function with multivariable and compact domain subset of R^d there exist a feed forward neural networks (FNNs) as an approximation for it . By a sigmoidal we can be approximated arbitrarily well .

A three-layer of the FNNs with d input units and one hidden and one output units can be mathematically expressed as

$$N_n(x) = \sum_{i=1}^{m} c_i \sigma\left(\langle \omega_{ij} x_j \rangle + \theta_i\right), \; x \epsilon R^d , d \geq 1 , \qquad (1.1)$$

where $1 \leq i \leq m$, $\theta_i \in R$ is the threshold, $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{id})^T \in R^d$ are connection weights of neuron i in the hidden layer with input neurons, $c_i \in R$ are the connection strength of neuron i with the output neuron, and σ is the sigmoidal activation function used in the network.

In this paper we prove direct and inverse estimation and saturation problem for the approximation of multivariate function in $L_p(K)$ spaces for $p < 1$ using a forward regular neural network .

2. Notations and Definitions

Let R be the set of reals , R^d be the d-dimensional Euclidean space ($d \geq 1$) , and let K be any subset of R^d

Definition 2.1

Let K be a multiple cell in d-dimension Euclidean space R^d ($d \geq 1$), **the L_p *space*** for $p < 1$ defined by :

$$L_p(K) = \{f: K \to R \,;\ \|f\|_p = (\int_k |f|^p)^{\frac{1}{p}} < \infty\}\,.$$

For any $x = (x_1, x_2, \dots, x_d)$ and $y = (y_1, y_2, \dots, y_d)$

Let $d(x, y)$ be the Euclidean distance of x and y , that is ,

$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_d - y_d)^2}$

Definition 2.2

the ***r − th order difference*** of function $f \in L_p(K)$

$$\Delta_h^r f(x) = (I - T^h)^r f(x) \quad , x = x_1, x_2, \dots, x_d \in K \quad , r \in N \text{ and } h > 0$$

we use I for the unit operator.

For any positive integer r we define the generalized ***module of smoothness of the rth order*** by the formula

$\omega_r(f, \delta)_p = \sup_{0<h<\delta} \|\Delta_h^r f\|_p\,, \quad \delta > 0, f \in L_p(K)\,.$

Definition 2.3

we denoted by the Lipschitzian class $Lip(\alpha)_r$ defined by the space of all functions f in $L_p(K)$ spaces satisfies $\omega_r(f, t)_p = O(t^\alpha)$, where $0 < \alpha \leq r$.

Remarks 2.4

1) In this paper we deal with the approximation by neural network with special type of neural activation functions J_k of which each function $\sigma: R \to [0,1]$ has up to $K + 1$ order continuous derivatives $\sigma^k, k = 1,2, \dots, K + 1$.

2) The regular neural activation functions are the normal sigmoidal activation functions $\sigma(x) = \frac{1}{1+e^{-\alpha x}}$ for positive .

3) Any neural network whose neural activation functions are regular (satisfies the conditions of part 1 of this remark) will be called a regular neural network.

Definition 2.5

Let r be an integer number and

$$P_r(x) = a_r x^r, x \in [a, b] \subset (-\infty, \infty) \qquad (2.6)$$

be a ***homogeneous univariate polynomial of degree r***.

XU Zongben & CAO Feilong (2004) prove the lemma "Let [a, b] be a compact interval, $\sigma \in J_r$ be a regular neural activation function and $P_r(x)$ a homogeneous univariate polynomial of the form (2.6). Then for any $\epsilon > 0$, there is a neural network of the form (1.1) the number of whose hidden units is not less than (r + 1) such that $|N_n(x) - P_r(x)| < \epsilon$ "

As a direct consequence of above lemma we introduce the following theorem .

Theorem 2.7

For any regular neural activation function $\sigma \in J_r$ and a homogeneous univariate polynomial $P_r(x)$ and a given $\epsilon > 0$ there is a neural network of the form (1.1) with not less than $r + 1$ hidden layers such that $\|N_n(x) - P_r(x)\|_p < \epsilon$

Proof:

Since $\sigma \in J_r$ be a regular neural activation function and $P_r(x)$ a homogeneous univariate polynomial then by above lemma we have for any $\epsilon > 0$ there is a neural network of the from (1.1) the number of whose hidden units is not less than $(r+1)$ such that :

$$|N_n(x) - P_r(x)| < \frac{\epsilon}{(\mu(k))^{1/P}}$$

$$\Rightarrow |N_n(x) - P_r(x)|^P < \left(\frac{\epsilon}{(\mu(k))^{1/p}}\right)^p$$

$$\Rightarrow \int_K |N_n(x) - P_r(x)|^P < \int_k \left(\frac{\epsilon}{(\mu(k))^{1/p}}\right)^p$$

$$\Rightarrow \left(\int_K |N_n(x) - P_r(x)|^P\right)^{1/P} < \left(\int_k \left(\frac{\epsilon}{(\mu(k))^{1/p}}\right)^p\right)^{1/p}$$

$$\Rightarrow ||N_n(x) - P_r(x)||_p < \frac{\epsilon\,(\mu(k)))^{1/p}}{(\mu(k))^{1/p}}$$

$$\Rightarrow ||N_n(x) - P_r(x)||_p < \epsilon$$

In this section we construct an FNN to realize universal approximation to any integral multivariate functions in $L_p(K)$, $P < 1$ we will use the Bernstein _ Durrmeger operation as base tools . ■

Definition 2.8 [X. Zongben and C. Feilong , (2004)]

Let K be any subset of R^d the Bernstein _Durremger operater B_n in $L_1(T)$ defined by :

$$(B_n f)(\alpha) = \sum_{|K| \le n} P_{n,k}(x)\ \phi_{n,k}(f) \qquad (2.9)$$

Where $x \in K$, $f \in L_1(K)$ and "

$$\phi_{n,k}(f) = (n+d)!/n! \int_k P_{n,k}(u)\, f(u) du$$

Lemma 2.10

If $f \in Lp\ (k)$ for $p < 1$ then

$$||B_n f - f||_p \le c(p)\ \omega_r(f, 1/n)_p$$

Proof: $||B_n f - f||_P \le c(p)\ K_r(f, (1/n)^r)_P$

$\le c(p)\ \omega_r(f, \frac{1}{n})_p$ [E. S. Belkina and S. S. Plato)ov (2008)]

Lemma 2.11

If $f \in Lp\ (k)$ for $p < 1$ then

$$\omega_r(f, 1/n) \le c(p) \sum_{i=1}^{n} ||B_i f - f||_p$$

Proof// $\omega_r(f,\delta)_p = \omega_r(f - B_n f + B_n f, \delta)_p \le c(p)\ \omega_r(f - B_n f, \delta) + \omega_r(B_n f, \delta) \le c(p)\ \left||f - B_n f|\right| + J$

$$B_n(f) - B_0(f) = B_n(f) - B_{2^l}(f) + \left(B_{2^l}(f) - B_{2^l-1}(f)\right) + \cdots + \left(B_1(f) - B_0(f)\right)$$

$2^l = n$, $l = \max i$, $2^i < n$

$J = \omega_r(B_n f, \delta) = \omega_r(\sum_{i=1}^{l} B_{2^i} - B_{2^{i-1}}, 2^{-l})_p$

$= \omega_r(\sum_{i=1}^{l} B_{2^i} - f + f - B_{2^{i-1}}, 2^{-l})_p$

$\le c(p) \sum_{i=1}^{l} ||f - B_{2^i} f||_p$

$\le c(p) \sum_{i=1}^{n} ||f - B_i f||_p$

$\therefore \omega_r(f,\delta) \le c(p) \sum_{i=1}^{n} ||B_i f - f||_p$. ■

To explain Lemma 2.13 we need the flowing notations: [X. Zongben and C. Feilong , (2004)]

1. Let z_+^d be the set of all non-negative multi-integers in R^d

2. For any $x = (x_1, x_2, \dots, x_d) \in R^d$ and $k = (k_1, k_2, \dots, k_d) \in Z_+^d$, Let $|x| = \sum_{i=1}^{d} x_i$, $|k| = \sum_{i=1}^{d} k_i$, $x^k = x_1^k \dots x_d^{kd}$ and $k! = k_1! \, k_2! \dots k_d!$

3. We say that $x \le y$, for any $, y \in R^s$, iff $x_i \le y_i$ for any $1 \le i \le s$.

4. For any fixed point p Let $N_p = \binom{p+d-1}{d-1}$

Be the number of multi-integers $\mathrm{i} = (\mathrm{i}_1, \mathrm{i}_2, \dots, \mathrm{i}_d)$in Z_+^d that satisfy $\mathrm{i}_1 + \mathrm{i}_2 + \dots + \mathrm{i}_d = \mathrm{p}$

5. Let $I_p = \binom{p+d-2}{d-2}$ be the number of multi-integers $j = (j_1, j_2, \dots, j_d)$ in Z_+^{d-1} that satisfy $j_1 + j_2 + \dots + j_{d-1} = p$

6. Denote by $j_{Np-1} + l, 1 \le l \le I_p$ a generic multi-integers j in Z_+^{d-1} satisfing $j_1 + j_2 + \dots + j_{d-1} = p$

7. j_l, $1 \le l \le N_p$ a generic multi-integers j in Z_+^{s-1} satisfing $j_1 + j_2 + \dots + j_{d-1} = p$

8. For any $1 \le l \le N_p$ Let $i_l^{(p)} = (p - |j_l|, j_l)$ then each $i_l^{(p)} = p$ is multi-integer in Z_+^d that satisfies $\left|i_l^{(p)}\right| = p$

9. Define $p_1 = (1, j_l), 1 \le l$ and $p_l^{(p)} = \frac{1}{2(1+p)} p_1, 1 \le l \le N_p$

we then have $\left|\mathrm{p}_l^{(p)}\right| \le \frac{1}{2}$ for any $1 \le \mathrm{l} \le \mathrm{N_p}$ "

The flowing lemma provides an equivalent expression of Bernstein-durrmeyer operator $\mathrm{B_n}$. ■

Lemma 2.12 [X. Zongben and C. Feilong , (2004) **]**

For any $\in L_p(T)$, the Berustin –Durrmeyer operator $B_n f$ in (2.9) can be expressed as :

$$B_n f(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} \langle x, p_l^{(n)} \rangle^p \qquad (2.13)$$

where $\langle x, p_l^{(n)} \rangle$ is inner product x and $p_l^{(n)}, d_l^{(p)}$ are uniquely determined by

$$\begin{pmatrix} (j_1)^{j_1} & (j_2)^{j_1} & \dots & (j_{Np})^{j_1} \\ (j_1)^{j_2} & (j_2)^{j_2} & \dots & (j_{Np})^{j_2} \\ (j_1)^{j_{Np}} & (j_2)^{j_{Np}} & \dots & (j_{Np})^{Np} \end{pmatrix} \begin{pmatrix} d_1^{(p)} \\ d_2^{(p)} \\ \vdots \\ d_{Np}^{(p)} \end{pmatrix} = \left(\frac{2(1+n)^p}{p!}\right) \begin{pmatrix} i_1^{(p)}! & c_1^{(p)}(f) \\ i_2^{(p)}! & c_2^{(p)}(f) \\ & \vdots \\ i_{Np}^{(p)}! & c_{Np}^{(p)}(f) \end{pmatrix}$$

With $c_l^{(p)}(f) = \frac{n!}{(n-p)!} \sum_{q \le i_1^{(p)}} \Phi_{n,q}(f) \frac{1}{q!\left(i_1^{(p)} - q\right)!} (-1)^{\left|i_1^{(p)} - q\right|}$,

$1 \le l \le N_p$.

Remark 2.14

We observe that in expression (2.14) each term $\langle x, p_l^{(n)} \rangle$ can be viewed as homogeneous uniriate polynomial of $\langle x, p_l^{(n)} \rangle$ with order p , and so by Theorem 2.8, it can be approximated arbitrarily well by a network of the form

$N_{p+1}(x) = \sum_{i=1}^{Np} c_{i,p} \, \sigma\left(\omega_{i,p} \langle x, p_l^{(n)} \rangle + \theta\right), K_p \ge p + 1$

Remark 2.15

1) As $B_n f(x)$ can be approximate f, the following neural networks

$$N_n(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} \sum_{i=1}^{k_p} c_{i,p} \sigma\left(\omega_{i,p} \langle x, p_l^{(n)} \rangle + \theta\right) \qquad (2.16)$$

Then can approximate f to any accuracy .
2) the network (2.17) will be the FNN models we propose use in this paper.
3) the network (2.17) are clearly of from (1.1) and contain hidden units where

$$m = N_0 + k_1N_1 + \cdots + k_nN_n$$
$$\geq N_0 + 2N_1 + \cdots + (n+1)N_n$$
$$= \sum_{k=0}^{n}(k+1)\binom{k+d-1}{d-1} = m_0(n) .$$

Theorem 2.17

For any $f \in L_p(K)$, $p < 1$ there is a regular, one hidden layer FNN, $N_n(x)$, of the form (1.1) with $\sigma \in J_n$ and the hidden unit number $m \geq \sum_{k=0}^{n}(k+1)\binom{k+d-1}{d-1} = m_0(n)$ such that

$\|N_n - f\|_p \leq c(p)\ \omega_r(f,\frac{1}{n})_p$

Proof : we assume $f \in L_p(K)$

Then, by Lemma 2.12, the Bernstein –durrmeyer operator B_nf can be defined and expressed as $B_nf(x) = \sum_{p=0}^{n}\sum_{l=1}^{N_p} d_l^{(p)}\langle x, p_l^{(n)}\rangle^p$

and, furthermore, it approximates f in the following sense

$\|B_nf - f\|_p \leq c(p)\ \omega_r(f,\frac{1}{n})_p$ [lemma 2.10]

And by remark (2.15) part 1 , we have

$\|N_n - f\|_p \leq c(p)\|B_nf - f\|_p \leq c(p)\ \omega_r(f,\frac{1}{n})_p$. ■

Theorem 2.18

For any $f \in L_p(K)$, $p < 1$ there is a regular, one hidden layer FNN, $N_n(x)$, of the form (1.1) with $\sigma \in J_n$ and the hidden unit number $m \geq \sum_{k=0}^{n}(k+1)\binom{k+d-1}{d-1} = m_0(n)$ such that

$c(p)\omega_r(f,\frac{1}{n})_p \leq \sum_{i=1}^{n}\|N_i - f\|_p$

Proof : we assume $f \in L_p(K)$

Then, by Lemma 2.12, the Bernstein –durrmeyer operator B_nf can be defined and expressed as $B_nf(x) = \sum_{p=0}^{n}\sum_{l=1}^{N_p} d_l^{(p)}\langle x, p_l^{(n)}\rangle^p$ (2.13)

and since $\left|p_l^{(n)}\right| \leq \frac{1}{2}$, we have $\langle x, p_l^{(n)}\rangle \leq 1$

i.e $(-1 \leq \langle x, p_l^{(n)}\rangle \leq 1)$

Each term $\langle x, p_l^{(n)}\rangle^p$ in (2.13) is univariate homogeneous polynomial of $\langle x, p_l^{(n)}\rangle$ with order p define on $[-1,1]$

Now by Theorem 2.7 we have $\langle x, p_l^{(n)}\rangle^p$ can be approximated by neural network

$$N_{K_p} = \sum_{i=1}^{Np} c_{p,i}\,\sigma\left(\omega_{p,i}\langle x, p_l^{(n)}\rangle + \theta\right) , c_{p,i}, \omega_{p,i} \in R , K_p \geq p+1 \quad (2.19)$$

With accuracy

$$\left\|N_{K_p} - \langle x, p_l^{(n)}\rangle^p\right\|_p \leq \epsilon \qquad (2.20)$$

Then for constructed FNN

$N_n(x) = \sum_{p=0}^{n}\sum_{l=1}^{N_p} d_l^{(p)} \sum_{i=1}^{kp} c_{i,p}\sigma\left(\omega_{i,p}\langle x, p_l^{(n)}\rangle + \theta\right)$, $c_{p,i}, \omega_{p,i} \in R$, $K_p \geq p+1$

and we have

$\|N_n - f\|_p = \|N_nf - B_nf + B_nf - f\|_p$

$$\leq \|N_n f - B_n f\|_p + \|B_n f - f\|_p$$
$$\leq c(p)\ \omega_r(f, \frac{1}{n})_p + \|N_n f - B_n f\|_p \qquad (2.21)$$

The term $\|N_n f - B_n f\|_p$ in above can be arbitrarily small, because (2.19) and (2.20) imply

$$\|N_n f - B_n f\|_p = \left\| \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} \{\langle x, p_l^{(n)}\rangle^p - \sum_{i=1}^{kp} c_{i,p} \sigma\left(\omega_{i,p}\langle x, p_l^{(n)}\rangle + \theta\right)\} \right\|_p$$
$$= \sum_{p=0}^{n} \sum_{l=1}^{N_p} \left|d_l^{(p)}\right| \max \left|\langle x, p_l^{(n)}\rangle^p - N_{K_p}\right| \leq \epsilon \sum_{p=0}^{n} \sum_{l=1}^{N_p} \left|d_l^{(p)}\right|$$

Then inequality (2.21) imply $\|B_n f - f\|_p \leq c(p)\ \omega_r(f, \frac{1}{n})_p$

and by (Lemma 2.11) $c(p)\ \omega_r(f, \frac{1}{n})_p \leq c(p) \sum_{i=1}^{n} ||B_i f - f||_p$

Then $c(p)\ \omega_r(f, \frac{1}{n})_p \leq \sum_{i=1}^{n} ||N_n f - f||_p$. ■

Theorem 2.23

For any $f \in L_p(K)$, $p < 1$ there is a regular, one hidden layer FNN, $N_n(x)$, of the form (1.1) with $\sigma \in J_n$ and the hidden unit number $m \geq \sum_{k=0}^{n} (k+1)\binom{k+d-1}{d-1} = m_0(n)$ such that

$\|N_n f - B_n f\|_p = O(n^{-\frac{\alpha}{2}})$ if and only if $f \in Lip(\alpha)_r$

Proof: let $f \in Lip(\alpha)_r$

From $\|N_n - f\|_p \leq c(p)\ \omega_r(f, \frac{1}{n})_p \Longrightarrow \|N_n - f\|_p = O(\frac{1}{n})^{\alpha}$

And by $c(p)\ \omega_r(f, \frac{1}{n})_p \leq \sum_{i=1}^{n} ||N_i f - f||_p$

We obtain $\omega_r(f, \frac{1}{n})_p = O\left(\frac{1}{n^{\alpha}}\right)$.■

References

Belkina E.S. and S. S. Platonov , (2008) , Equivalence of K-Functionals and Modulus of Smoothness Constructed by Generalized Dunkl Translations , ISSN 1066-369X, Russian Mathematics (Iz. VUZ), Vol. 52, No. 8, pp. 1–11.

Cardaliaguet, P., Euvrard, G., (1992) ‹ Approximation of a function and its derivatives with a neural network, Neural Networks , 5: 207—220.

Chen, T. P., (1994),Approximation problems in system identification with neural networks, Science in China, Series A, , 24(1): 1—7.

Chen, T. P., Chen, H., (1995) , Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical system, IEEE Transaction on Neural Networks, , No. 6,pp. 911—917.

Chui, C. K., Li, X., (1992), Approximation by ridge functions and neural networks with one hidden layer, J. Approx. Theory, , 70: 131—141.

Chui, C. K., Li, X., (1993) , Neural networks with one hidden layer, In: Multivariate Approximation: From CAGD to Wavelets (K. Jetter, F. I. Utreras, eds), pp.77—89, Singapore: World Scientific Press .

Cybenko, G.,(1989), Approximation by superpositions of sigmoidal function, Math.of Control Signals, and System, , 2: 303—314.

Funahashi, K. I.,(1989), On the approximate realization of continuous mappings by neural networks, Neural Networks, , 2: 183—192.

Gallant, A. R., White,(1992) , H., On learning the derivatives of an unknown mapping with multilayer feedforward networks, Neural Networks, , 5: 129—138.

Hornik, K., Stinchombe, M., White, H.,(1990) , Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Networks, , 3: 551—560.

Hornik, K., Stinchombe,(1989) , M., White, H., Multilayer feedforward networks are universal approximation,

Leshno, M., Lin, V. Y., Pinks, A.,(1993), Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Networks, , 6: 861—867.

Mhaskar, H. N., Micchelli, C. A.,(1992), Approximation by superposition of a sigmoidal function, Advances in Applied Mathematics, , 13: 350—373.

Neural Networks, , 2: 359—366.

Zongben X. and C. Feilong , (2004) , The essential order of approximation for neural networks , Science in China Ser. F Information Sciences , Vol.47 No.1 97—112 .