Journal of Babylon University, Pure and Applied Sciences, No.(1), Vol.(26): 2018

Nearly Exponential Neural Networks Approximation in L_p Spaces

Eman Samir Bhaya **Zahraa Mahmoud Fadel**
Mathematics Department, College of Education for Pure Sciences, University of Babylon.
emanbhaya@itnet.uobabylon.edu.iq 12345mathzahra@gmail.com

Abstract
In different applications, we can widely use the neural network approximation. They are being applied to solve many problems in computer science, engineering, physics, etc. The reason for successful application of neural network approximation is the neural network ability to approximate arbitrary function. In the last 30 years, many papers have been published showing that we can approximate any continuous function defined on a compact subset of the Euclidean spaces of dimensions greater than 1, uniformly using a neural network with one hidden layer. Here we prove that any real function in L_P (C) defined on a compact and convex subset C of $\mathbb{R}^d$can be approximated by a sigmoidal neural network with one hidden layer, that we call nearly exponential approximation.

Keywords: Nearly exponent function. Best approximation. Modulus of smoothness. Neural network approximation.

الخلاصة
يدخل التقريب باستخدام الشبكات العصبية في الكثير من التطبيقات المهمة. حيث انه يحل الكثير من المشاكل في مجالات علوم الحاسوب و الهندسة و الفيزياء، الخ. ان سبب نجاح التقريب باستخدام الشبكات العصبية هو امكانيته من تقريب اية دالة مهما كان نوعها. في الثلاثين سنة الماضية نشرت الكثير من البحوث جميع تلك البحوث بينت ان كل دالة معرفة على مجموعة مرصوصة محدبة و جزئية من الفضاء الاقليدي $\mathbb{R}^d$ يمكن تقريبها بانتظام باستخدام الشبكة العصبية ذات الطبقة المخفية الواحدة. في هذا البحث برهنا أن لأية داله تنتمي الى $L_p(C)$ و معرفه على مجموعه محدبة ومرصوصة C في $\mathbb{R}^d$ يمكن تقريبها باستخدام شبكه عصبيه ذات طبقه مخفيه واحده من نوع الاس القريب وهذا ما نسميه بالتقريب باستخدام الشبكات العصبية من نوع الاس القريب.

الكلمات المفتاحية : دالة الآس القريب. أفضل تقريب. معامل النعومة. التقريب باستخدام الشبكات العصبية.

1 Introduction and Basics
Artificial forward neural networks are nonlinear parametric expressions representing multivariate numerical functions. In connection with such paradigms there arise mainly three problems: a *density* problem, a *complexity* problem, and an *algorithmic* problem. The density *problem* deals with the following question: which functions can be approximated and, in particular, can all members of a certain class of functions be approximated in a suitable sense. This problem was satisfactorily solved in the late 1980’s [see (Cybenko,1989; Funahashi,1989; Hornik *et al.*,1989)]. Any continuous function on any compact subset of $\mathbb{R}^d$ can be uniformly approximated arbitrarily closely by a neural network with one hidden layer. Moreover, the proof given in (Hornik *et al.*,1989) provides an intimate connection forward neural networks and polynomials. (Ratter,1999)
In this paper we improve the works in (Cybenko,1989; Funahashi, 1989; Hornik *et al.*, 1989, Ranjita R., 2018) and introduce a direct theorem using neural weights in term of polynomial approximation of functions in L_P spaces .

Let N be the set of nonnegative integer numbers, and let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}^+$ be the set of nonnegative real numbers, $\mathbb{R}^d$ be the d-dimensional Euclidean space $(d \geq 1)$, and let Λ be a finite space subset of $\mathbb{R}^+$, let $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $y = (y_1, , \dots, y_d) \in \mathbb{R}^d$, $e^x = (e^{x_1}, e^{x_2}, \dots, e^{x_d})$, $x^y = ({x_1}^{y_1}, {x_2}^{y_2}, \dots, {x_d}^{y_d})$, $x_j \geq 0, j = 1,2,3, \dots, d$. and let $\mathbb{p}_n(d)$ be the space of all d_variate algebraic polynomial, also we use the *active* function $\delta: R \to R$ is *nearly exponential* .
Let f be a real valued function defined on a convex subset $C \subset \mathbb{R}^d$. Define

$$L_P(C) = \left\{ f: C \to \mathbb{R} \ : \ \|f\|_{L_P(C)} = \underbrace{\int_C \dots \int_C}_{d\ times} (|f|^P)^{\frac{1}{P}} < \infty \right\}.$$

For any $f \in L_P(C)$ and a real or complex function set S , *the distance* from f to S defined by : $d_P(f, S) = sup_{g \in S} \|f - g\|_{L_P(C)}$.
The *rth symmetric difference* of f is given by

$$\Delta_h^r(f, r, C) = \Delta_h^r(f, x) = \begin{cases} \sum_{i=0}^{r} (-1)^{r-i} \binom{r}{i} f\left(\left(x_1 - \frac{rh}{2} + ih\right), \left(x_2 - \frac{rh}{2} + ih\right), \dots, \left(x_d - \frac{rh}{2} + ih\right) \right) & x \pm \frac{rh}{2} \in C \\ 0 & o.w \end{cases}$$

Then the *rth usual modulus of smoothness* of $f \in L_P(C)$ is defined by

$$\omega_r(f, \delta, C)_{L_P(C)} \coloneqq sup_{|h| \leq \delta} \|\Delta_h^r(f, .)\|_{L_P(C)}\ \delta \geq 0. \qquad \text{(Bhaya, 2003)}$$

Now let us recall the mathematical expression of the neural network, a three-layer of FNN with one hidden layer, d inputs and one output can be mathematically expressed as $N(x) = \sum_{i=1}^{m} c_i\, \sigma\left(\sum_{j=1}^{d} w_{ij}\, x_j + \theta_i\right), x \in \mathbb{R}^d, d \geq 1$, where $1 \leq i \leq m, \theta_i \in \mathbb{R}$ are the thresholds, $w_i = (w_{i1}, w_{i2}, \dots, w_{id}) \in \mathbb{R}^d$ are connection weights of neuron i in the hidden layer, with the hidden layer of the input neurons, $c_i \in \mathbb{R}$ are the connection strength of neuron i with the output neuron, and σ is the activation function used in the network. (Wang and Zonghen, 2010)

$f \in L_P(C)$, is called *Lipchitz continuous*. If there exists $L > 0, h > 0$ such that $\|\Delta_h^1 f\|_{L_P(C)} \leq L(f)h$, then we get $L(f) = sup \frac{\|\Delta_h^1 f\|_{L_P(C)}}{h}$. An *exponential polynomial* of maximal degree $n \in N$ is of the form $\sum_{\propto \in \eta(0 \dots n)} a_\propto e^{-\propto x}$ for some $\propto > 0$. (Ritter, 1999) the symbol $P_n^E(d)$ stands for the set of all real, d-variate exponential polynomial of maximal degree n and arbitrary $\propto^-$. A function $\delta: R \to R$ is said to *nearly exponential* , whenever for any $\in > 0$, there exist real numbers ɤ, Ʈ, ß , ρ such that. $|ɤ\delta(Ʈt + ß) + \rho - e^t \ | < \in$, for all $t \leq 0$.
Given some activation function $\delta: R \to R$, $R_n^\delta(d)$ will be denote the set of all sums of the form $\sum_{ג \in \Lambda} \pm a\, \delta(-ג.x + b_ג)$. (Ritter, 1999) with $\Lambda \subseteq \eta(0 \dots \dots n)^d$ for some ɳ>0 and with a>0 independent of ג .
From now on we shall use the notation $C(p, r, d)$ for the absolute constant depending on p , r , d only and not the same for all steps in our proofs.
As an auxiliary result we need the following theorem from. (Kareem, 2011).
Theorem 1.1 (Kareem, 2011)
If $f \in L_P[a,b]^d, 0 < P < \infty$ then $E_{n-1}(f)_P \leq C(p, m, d)\ \omega_m(f, h, [a, b]^d)_{L_P[a,b]^d}$ where $E_{n-1}(f)$ is the degree of best approximation of f by algebraic polynomial of degree $\leq n - 1$, which is $E_{n-1}(f) = \inf_{f \in \mathbb{p}_{n-1}} \|f - P_{n-1}\|_{L_P[a,b]^d}$, $\mathbb{p}_n$ is the space of all algebraic polynomials of degree $\leq n$.

2. The Main Results

Here, let us introduce our main results.

Theorem 2.1

For any $f \in L_p[0,1]^d$, we have $d_P\left(f, P_n^E(d)\right) \le C(p,r,d)\, \omega_r(f, \frac{1}{n})_{L_P[0,1]^d}$

Proof

Use Theorem1.1 to approximate the function $f \in L_p[0,1]^d$ by an algebraic polynomial of the form $P(x) = \sum_{\alpha \in (0\ldots n)^d} a_\alpha \prod_{i=1}^d x_i^{\alpha_i}$ and satisfy

$$\|P-f\|_{L_P} \le C(p,r,d)\, \omega_r\left(f, \frac{1}{n}\right)_{L_P[0,1]^d}.$$

The sequence $\langle F_\mu(x)\rangle, \mu \in IR^+$ converges to the identity function F_0. choose μ such that for a given $\in > 0$

$$\left\|P(F_\mu) - P(F_0)\right\|_{L_P[0,1]^d} < \in \quad (1)$$

It is clear, $P(F_\mu)$ is an exponential polynomial $P(F_\mu)$ in $P_n^E(d)$, and

$$d_P\left(f, P_n^E(d)\right) \le \left\|P(F_\mu) - f\right\|_{L_P[0,1]^d} \le \left\|P(F_\mu) - P(F_0) + P(F_0) - f\right\|_{L_P[0,1]^d}$$

$$\le C(p)\left(\left\|P(F_\mu) - P(F_0)\right\|_{L_P[0,1]^d} + \|P(F_0) - f\|_{L_P[0,1]^d}\right) \quad (2)$$

Using (1) and (2) we get

$$d_P\left(f, P_n^E(d)\right) \le C(p,r,d)\, \omega_r\left(f, \frac{1}{n}\right)_{L_P[0,1]^d} + \in \quad (3)$$

Since (3) is true for any $\in > 0$, we get

$$d_P\left(f, P_n^E(d)\right) \le C(p,r,d)\, \omega_r\left(f, \frac{1}{n}\right)_{L_P[0,1]^d} \blacksquare$$

Theorem 2.2

For any $f \in L_P(C)$, we have $d_p\left(f, R_n^\delta(d)\right) \le C(p,r,d)\, \omega_r(f, \frac{1}{n})_{L_P(C)}$, where δ is nearly exponential and C is a compact and convex set in $[0,1]^d$.

Proof

Let ɭ denote the Euclidean projection $[\mathbf{0},\mathbf{1}]^d \to \boldsymbol{C}$, the function $\boldsymbol{f}$(ɭ) $\in L_P[0,1]^d$. Using Theorem 2.1 to approximate the function f(ɭ) by an exponential polynomial of the from $P(X) = \sum_{\propto \in ɳ(0\ldots\ldots n)^d} a_\propto e^{-\propto x}$, such that $\|P-f\|_{L_P(C)} \le \|P-f\|_{L_P[0,1]^d} \le C(p,r,d)\, \omega_r\left(f, \frac{1}{n}\right)_{L_P[0,1]^d} + \in$, let $\Lambda = \{\alpha \in ɳ(0 \ldots\ldots n)^d : a_\propto \ne 0\}$. Anther representation of P is $P(X) = \sum_{\propto\in\Lambda} \pm a\, e^{-\propto x + b_\alpha}$, where $b_\alpha \le 0$ and $a > 0$ independent of $\propto$. Since δ is nearly exponential function, since e^t is exponential function then we can approximate e^t by an expression ɤ δ(ƮBt+ ß)+ρ uniformly on the negative half line up to the error ∈. The sum

$S = \sum_{\Lambda\setminus\{0\}} \pm a\, ɤ\delta(-(\propto.\mathrm{x} + (b_\propto + ß)) + (\pm a\, ɤ\, \delta(b_0 + ß) + \rho \sum_{\propto\in\Lambda} \pm a)$.

Then $S \in R_n^\delta(d)$, $\|S-f\|_{L_P(C)} \le C(p,r,d)\, \omega_r\left(f, \frac{1}{n}\right)_{L_P(C)} + 2 \in$.

Similarly if $0 \in \Lambda$. Then we get

$$d_p\left(f, R_n^\delta(d)\right) \le C(p,r,d)\, \omega_r(f, \frac{1}{n})_{L_P(C)} \blacksquare$$

As a direct consequences of the above theorem we have the following corollaries.

Corollary 2.3
For any $f \in L_P(C)$ and any $\in > 0$, there exists a neural network of the form $N(x) = \sum \pm a\, \delta(-\lambda.x + b_\lambda)$ with at most $\min\{(n+1)^d / C(p,r,d)\ \omega_r \left(f, \frac{1}{n}\right)_{L_P(C)} < \in\}$ hidden neurons satisfy $\|N(x) - f\|_{L_P(C)} < \in$,where δ is nearly exponential and C is a compact and convex subset of $[0,1]^d$.

Proof
Choose n such that $C(p,r,d)\ \omega_r(f, \frac{1}{n})_{L_P(C)} < \in$. From Theorem 2.2 there exists $g \in R_n^\delta(d)$ Such that $d_P(g, f) < \in$. this g is neural network of hidden neurons$\le (n+1)^d$ ■

Corollary 2.4
Let δ be nearly exponential function and let C be a compact and convex set in $[0,1]^d$, f is a Lipchitz continuous function, then $\max\{[C(p,r,d)\ \frac{L(f)}{\in}]^d, 1\}$ neurons suffice.

Proof
We have two cases, the first case$C(p,r,d)\ \frac{L(f)}{\in} < 1$, then $C(p,r,d)\omega_r\left(f, \frac{1}{n}\right)_{L_P(C)} \le C(p,r,d)\frac{L(f)}{n+2} < \frac{\in}{n+2} < \in$. From Theorem 2.3 the minimum is equal to 1 . Therefore $C(p,r,d)\ \frac{L(f)}{\in} \ge 1$, let $\text{n} = [C(p,r,d)\ \frac{L(f)}{\in}] - 1$, then using Theorem2.3 the minimum is at most $(\text{n}+1)^d = [C(p,r,d)\ \frac{L(f)}{\in}]^d$ ■

Conclusions
We generalize the results proved in (Ranjita R., (2018)), for continuous functions to functions in L_p spaces. We prove that any function in L_p spaces defined on a convex and compact set can be nearly exponential approximated using sigmoidal neural network. This application makes the computer science researchers approximate any waves (the target function) of a neural network.

References

Bhaya E.S., (2003)," On the constrained and unconstrained approximation ", Ph. D. thesis, University of Baghdad, lraq.

Cybenko G., (1989), “Approximation by superpositions of a sigmoidal function,” Springer-Verlag New York inc. Mathematics of Control Signals and System., VOL.2, pp. 303–314.

Funahashi K.I., (1989), “On the approximate realization of continuous mappings by neural networks,”Science Direct-Neural Networks Journal, VOL. 2, pp. 183–192.

Hornik K., Stinchcombe M., and White H., (2011) “Multilayer-feed forward networks are universal approximators,” ACMDL Digital Librarym Journal of Neural Networks, VOL. 2, pp. 359–366.

Kareem M.A.,(2011), “On the multi approximation in suitable spaces ,” M.Sc. thesis, University of Babylon. Iraq.

Ranjita R., (2018), " A study of neural network based function approximation " International Journal of Engineering and Science Invention. " VOL. 7, NO. 1, pp. 32-34.

Ritter G., (1999), "Efficient estimation of neural weights by polynomial approximation," IEEE Transactions on Information Theory, VOL. 45, NO. 5, July , pp.1541-1550..

Wang J. and Zongben X., (2010), "New study on neural networks: the essential order of approximation," Science Direct, Neural Networks 23, pp. 618-624.