Nearly Exponential Approximation for Neural Networks

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 this paper we prove that any real function in $L_P(C)$ defined on a compact and convex subset of $\mathbb{R}^d$can be approximated by a sigmoidal neural network with one hidden layer , that we call nearly exponential.

Key words: Nearly exponent. Best approximation. Modulus of smoothness.

الخلاصة

برهنا في هذا البحث أن لأية داله معرفه على مجموعه محدبة ومرصوصة فيC $\mathbb{R}^d$ يمكن تقريبها باستخدام شبكه عصبيه ذات طبقه مخفيه واحده فقط في($L_P(C)$ وهذا ما نسميه بالتقريب الاسي الجانبي .

الكلمات المفتاحية : الآسي التقريبي . أفضل تقريب. معامل النعومة .

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 (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) and introduce a direct theorem using neural weights in term of polynomial approximation of functions in L_P spaces .

Let Nbe 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 $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $y = (y_1, , \dots, y_d) \epsilon \mathbb{R}^d$, $e^x = (e^{x_1}, e^{x_2}, \dots, e^{x_d})$, $x^y = ({x_1}^{y_1}, {x_2}^{y_2}, \dots, {x_d}^{y_d})$, $x_j \geq 0, j = 1,2,3, \dots, d$. 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 $C \in \mathbb{R}^d$. Define

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

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

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

Than *the rth usual modulus* of smoothness of $f\in L_P(C)$ is defined by
$\omega_r(f,\delta,C)_{L_P(C)}:=sup_{|h|\le\delta}\|\Delta_h^r(f,.)\|_{L_P(C)}\ \delta\ge 0.$ (Bhaya, 2003)

Now let us recall the mathematical expression the neural network, a three-layer of FNN with one hidden layer, d inputs and one output can be mathematically expressed as $N(x)=\sum_{i=1}^{m}c_i\,\sigma\left(\sum_{j=1}^{d}w_{ij}\,x_j+\theta_i\right), x\in\mathbb{R}^d, d\ge 1$ where $1\le i\le m, \theta_i\in\mathbb{R}$ 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 with 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 $\|\Delta_h^1 f\|_{L_P(C)}\le L(f)h$, than 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 $\sum_{\propto\in ▯(0\dots n)}a_\propto e^{-\propto x}$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 all $\in>0$, the exist real numbers ɤ, ▯, ß , ρ such that.

$$|▯\delta(▯t+ß)+\rho-e^t\ |<\in\quad,\qquad\text{for all } t\le 0$$

Given some activation function $\delta:R\to R$, $R_n^\delta(d)$ will be denote the set of all sums of the form $\sum_{ƛ\in\Lambda}\pm a\,\delta\left(-ƛ.x+b_ƛ\right)$. (Ritter, 1999) With $\Lambda\subseteq ▯(0\dots\dots n)^d$ *for* some ꞃ>0 and with a>0 independent of ƛ .

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 following theorem from .(Kareem, 2011)

Theorem 1.1 (Kareem, 2011)
If $f\in L_P[a,b]^d, 0<P<\infty$ then $E_{n-1}(f)_P\le C(p,m,d)\ \omega_m(f,h,[a,b]^d)_{L_P[a,b]^d}$
Where $E_{n-1}(f)$ is the degree of best approximation of $f\in L_P[a,b]^d$ by algebraic polynomial of degree $\le n-1$, 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 polynomial of degree $\le n$.

2. The Main Results

In this section we shall introduce our main results.

Theorem 2.1

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

Proof

Using Theorem 1.1 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

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

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

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

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

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

Using (1) and (2) we get

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

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

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

Theorem 2.2

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

Proof

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

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

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

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

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

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 set in $[0,1]^d$.

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$\leq (n+1)^d$. ■

Corollary 2.4
Let δ be nearly exponential function and let C be compact and convex set in $[0,1]^d$, f is 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 $C(p,r,d)\omega_r\left(f\,,\frac{1}{n}\right)_{L_P(C)} \leq C(p,r,d)\frac{L(f)}{n+2} < \frac{\in}{n+2} < \in$. From Theorem 2.3 the minimum is equal to 1 . therefore $C(p,r,d)\ \frac{L(f)}{\in} \geq 1$,let $\mathrm{n} = [C(p,r,d)\ \frac{L(f)}{\in}] - 1$, then using Theorem2.3 the minimum is at most
$(\mathrm{n}+1)^d = [C(p,r,d)\ \frac{L(f)}{\in}]^d$. ■

References

Bhaya E.S., (2003)," on the constrained and unconstrained approximation ",University of Baghdad ,lraq.

Cybenko G., (1989), “Approximation by superpositions of a sigmoidal function,” *Math. Contr. Signals Syst.*, vol. 2, pp. 303–314.

Funahashi K.I., (1989), “On the the approximate realization of continuous mappings by neural networks,” *Neural Networks*, vol. 2, pp. 183–192.

Hornik K. , Stinchcombe M. , and White H., (2011), “ Multilayer-feed forward networks are universal approximators,” Neural Networks, vol. 2, pp. 359–366, 1989.

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

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 in chin , Neural Networks 23 618_624a.