ARTICLE IN PRESS YACHA:1131

Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Provable approximation properties for deep neural networks

Uri Shaham [a,*], Alexander Cloninger [b], Ronald R. Coifman [b]

[a] *Statistics department, Yale University, United States*
[b] *Applied Mathematics program, Yale University, United States*

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 28 March 2016
Accepted 7 April 2016
Available online xxxx
Communicated by Charles K. Chui

Keywords:
Neural nets
Function approximation
Wavelets

ABSTRACT

We discuss approximation of functions using deep neural nets. Given a function f on a d-dimensional manifold $\Gamma \subset \mathbb{R}^m$, we construct a sparsely-connected depth-4 neural network and bound its error in approximating f. The size of the network depends on dimension and curvature of the manifold Γ, the complexity of f, in terms of its wavelet description, and only weakly on the ambient dimension m. Essentially, our network computes wavelet functions, which are computed from Rectified Linear Units (ReLU).

1. Introduction

In the last decade, deep learning algorithms achieved unprecedented success and state-of-the-art results in various machine learning and artificial intelligence tasks, most notably image recognition, speech recognition, text analysis and Natural Language Processing [12]. Deep Neural Networks (DNNs) are general in the sense of their mechanism for learning features of the data. Nevertheless, in numerous cases, results obtained with DNNs outperformed previous state-of-the-art methods, often requiring significant domain knowledge, manifested in hand-crafted features.

Despite the great success of DNNs in many practical applications, the theoretical framework of DNNs is still lacking; along with some decades-old well-known results, developing aspects of such theoretical framework are the focus of much recent academic attention. In particular, some interesting topics are (1) specification of the network topology (i.e., depth, layer sizes), given a target function, in order to obtain certain approximation properties, (2) estimating the amount of training data needed in order to generalize the test data with high accuracy, and also (3) development of training algorithms with performance guarantees.

* Corresponding author.
E-mail address: uri.shaham@yale.edu (U. Shaham).

http://dx.doi.org/10.1016/j.acha.2016.04.003
1063-5203/

1.1. The contribution of this work

In this manuscript we discuss the first topic. Specifically, we prove a formal version of the following result:

Theorem (informal version) 1.1. *Let $\Gamma \subset \mathbb{R}^m$ be a smooth d-dimensional manifold, $f \in L_2(\Gamma)$ and let $\delta > 0$ be an approximation level. Then there exists a depth-4 sparsely-connected neural network with N units where $N = N(\delta, \Gamma, f, m)$, computing the function f_N such that*

$$\|f - f_N\|_2^2 \leq \delta. \tag{1}$$

The number $N = N(\delta, \Gamma, f, m)$ depends on the complexity of f, in terms of its wavelet representation, the curvature and dimension of the manifold Γ and only weakly on the ambient dimension m, thus taking advantage of the possibility that $d \ll m$, which seems to be realistic in many practical applications. Moreover, we specify the exact topology of such network, and show how it depends on the curvature of Γ, the complexity of f, and the dimensions d, and m. Lastly, for two classes of functions we also provide approximation error rates: L_2 error rate for functions with sparse wavelet expansion and point-wise error rate for functions in C^2:

- if f has wavelet coefficients in l_1 then there exists a depth-4 network and a constant c so that

$$\|f - f_N\|_2^2 \leq \frac{c}{N} \tag{2}$$

- if $f \in C^2$ and has bounded Hessian, then there exists a depth-4 network so that

$$\|f - f_N\|_\infty = O\left(N^{-\frac{2}{d}}\right). \tag{3}$$

1.2. The structure of this manuscript

The structure of this manuscript is as follows. In Section 2 we review some of the fundamental theoretical results in neural network analysis, as well as some of the recent theoretical developments. In Section 3 we give quick technical review of the mathematical methods and results that are used in our construction. In Section 4 we describe our main result, namely construction of deep neural nets for approximating functions on smooth manifolds. In Section 5 we specify the size of the network needed to learn a function f, in view of the construction of the previous section. Section 6 concludes this manuscript.

1.3. Notation

Γ denotes a d-dimensional manifold in $\mathbb{R}^m$. $\{(U_i, \phi_i)\}$ denotes an atlas for Γ. Tangent hyper-planes to Γ are denoted by H_i. f and variants of it stand for the function to be approximated. Furthermore, φ, ψ are scaling (aka "father") and wavelet (aka "mother") functions, respectively. The wavelet terms are indexed by scale k and offset b. The support of a function f is denoted by $\text{supp}(f)$.

2. Related work

There is a huge body of theoretical work in neural network research. In this section, we review some classical theoretical results on neural network theory, and discuss several recent theoretical works.

A well-known result, proved independently by Cybenko [5], Hornik [10] and others states that Artificial Neural Networks (ANNs) with a single hidden layer of sigmoidal functions can approximate arbitrary closely any compactly supported continuous function. This result is known as the "Universal Approximation Property". It does not relate, however, the number of hidden units and the approximation accuracy;

Please cite this article in press as: U. Shaham et al., Provable approximation properties for deep neural networks, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.04.003