Approximation by Superpositions of a Sigmoid Function

Biplab Kumar Pradhan, 13602

April 21, 2017

Abstract

This report surveys the proof of uniform approximation of continous functions by sigmoid superpositions [1]. The proof implies that any multivariate continous function in the unit hypercube $[0, 1]^m$ can be approximated by a finite sum of univariate sigmoid functions with suitable parameters. This series can be represented as a neural network with one hidden layer, sigmoid being the activation function of each unit, and a linear output layer.

1 Introduction

A sigmoid function is given by

$$\sigma(z) = \frac{1}{1 + e^{(-z)}}$$

with $z \in R^1$

Let $F(x)$ be a continous function of m variables in $[0, 1]^m$. It can be proven that [1] , the following finite series can approximate this function upto any desired precision.

$$\sum_{i=1}^{N} v_i \sigma(w_i^T x + b_i)$$

This is known as the Universal Approximation Theorem.

The sigmoid function is of particular significance to neural network theory, where it's used as an activation function for its nodes. A neural network is a computational model whose parameters are trained on a given dataset, to be able to predict accurate output for inputs not in the dataset. Its use is prominent in pattern classification [2].

With the Universal Approximation Theorem, it's proven that any continous function in a compact domain can be uniformly approximated by a feed-forward neural network with finite number of neurons.

2 An Informal Proof

Let us just consider the 1-D domain $[0,1]$ for simplicity. We can intuitively justify that the sigmoidal series can uniformly approximate any function [3], by considering that:

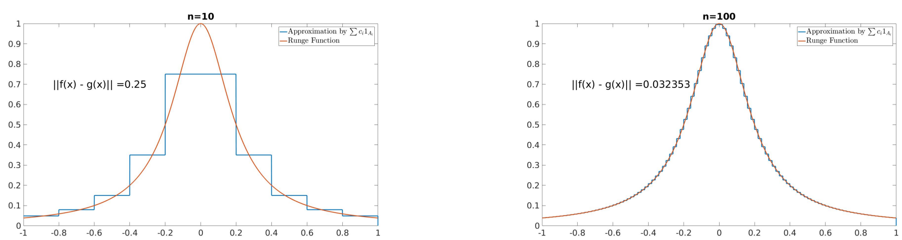
- Any function can be uniformly approximated by a piecewise constant function.
- A piecewise constant function can be written as a series of indicator functions of the form

$$G(x) = \sum_{i=1}^{N} c_i 1_{A_i}(x)$$

where $A_i \subset [0,1]$ is the i^{th} subinterval in $[0,1]$ after it's divided into N panels. The optimum value of each c_i can be easily computed by

$$c_i = \frac{\max_{x \in A_i} f(x) + \min_{y \in A_i} f(y)}{2}$$

- An Indicator function can be approximated to arbitrary precision by a sigmoid function by manipulating its slope.

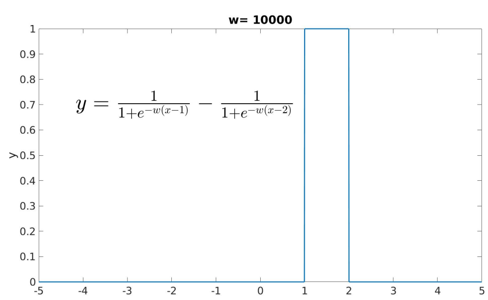


(a) 10 Indicator functions

(b) 100 Indicator functions

Figure 1: Approximation by series of Indicator functions

Figure 2: Two sigmoid functions used to approximate an indicator function of $[1,2]$



3 Cybenko's Proof [1]

Let I_n denote our domain, $[0,1]^n$, and $C(I_n)$ a function space of all continous functions on $[0,1]^n$. Let $M(I_n)$ denote the space of all regular signed Borel measures. A measure is a real-valued function defined on a set. It maps subsets to real-values. A Borel measure is a measure defined on a Borel set.

Definition: A discriminatory function is defined by:

$$\int_{I_n} \sigma(y^T x + b) d\mu \neq 0$$

for any $y \in R^n$ and $b \in R$.

Lemma 1: The sigmoid function is discriminatory. The proof of this Lemma can be found in [1].

Lemma 2: The function space $C(I_n)$ is an closed set.

Proof: Let $B(I_n)$ be the set of all bounded functions defined on I_n. Then, $C(I_n) \subset B(I_n)$. To prove that $C(I_n)$ is closed, we need to show that all its limit points exist inside the set.

Let $f \in B(I_n)$ be a limit point of $C(I_n)$, i.e. there exists a sequence of functions $\{f_n\} \in C(I_n)$ that converge to f. For any arbitrary ϵ we can pick a $f_N \in \{f_n\}$ such that,

$$||f - f_N|| < \epsilon/3$$

As f_N is continous, we can pick a δ such that $|f_N(x) - f_N(y)| < \epsilon/3$ for $|x - y| < \delta$.

Now, for $|x - y| < \delta$

$|f(x) - f(y)| = |f(x) - f_N(x) + f_N(x) - f_N(y) + f_N(y) - f_N(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_N(y)|$
$\implies |f(x) - f(y)| < \epsilon/3 + \epsilon/3 + \epsilon/3$
$\implies |f(x) - f(y)| < \epsilon$

Thus, for any $\epsilon > 0$ we can pick a $\delta > 0$ for which the above inequality holds. Therefore f is continous $\implies f \in C(I_n)$.

Thus, $C(I_n)$ has all its limit points, and therefore it's a closed set.

Lemma 3: Let $S \subset C(I_n)$ be the space of all functions of the form

$$G(x) = \sum_{i=1}^{N} v_i \sigma(w_i^T x + b_i)$$

With σ being the sigmoid function. Then S is an open set.

Proof: Let us consider the function

$$g(x) = \sigma(wx)$$

We can see that $g \in S$. Let $w \to \infty$, then $g \to 1$, but $1 \notin S$. Therefore, S does not have all its limit points, and hence it's an open set.

3.1 Main Proof

Our approach to the proof involves proving that the $closure(S) = C(I_n)$. This would imply that for every $f \in C(I_n)$ there is a sequence $\{f_n\} \in S$ such that, there is an N for every $\epsilon > 0$ for which $||f_N - f||_\infty < \epsilon$.

We will proceed by contradiction. Let us assume that $closure(S) \neq C(I_n)$. Let $closure(S) = R$ for some $R \subset C(I_n)$. By Hahn-Banach Theorem there exists some bounded linear mapping L such that $L(R) = L(S) = 0$ and $L \neq 0$.

Riesz Representation Theorem implies that,

$$L(h) = \int_{I_n} h(x) d\mu(x)$$

for some $\mu \in M(I_n)$, for all $h \in C(I_n)$. As $\sigma(y^T x + b) \in R \subset C(I_n) \implies$

$$\int_{I_n} \sigma(y^T x + b) d\mu(x) = 0$$

for all y and b.

By Lemma 1, we know that that this integral will never be 0. Thus, we have a contradiction, and $closure(S) = C(I_n)$, and our theorem is proved.

References

[1] Cybenko, George. "Approximation by superpositions of a sigmoidal function." Mathematics of Control, Signals, and Systems (MCSS) 2.4 (1989): 303-314. APA

[2] Lippmann, Richard. "An introduction to computing with neural nets." IEEE Assp magazine 4.2 (1987): 4-22. APA

[3] Nielsen, Michael. "Visual proof that neural nets can compute any function." http://neuralnetworksanddeeplearning.com/chap4.html