Approximation by Superpositions of a Sigmoidal Function

M. Ruby, C. Salinas

Department of Mathematics
Purdue University

Purdue University Machine Learning Seminar, 2019

Introduction

We wish to approximate a function $f : \mathbf{R}^n \to \mathbf{R}$ by an expression of the form

$$\sum_{j=1}^{N} \alpha_j \sigma(\mathbf{y}_j^{\mathsf{T}} \mathbf{x} + \theta_j), \tag{1}$$

where $\mathbf{y}_j \in \mathbf{R}^n$ and $\alpha_j, \theta_j \in \mathbf{R}$ are fixed. This is exactly a neural network with one hidden layer.

The Main Result

The main result is the following:

Theorem

For σ a continuous sigmoidal function, finite sums of the form in Equation (1), i.e.,

$$\sum_{j=1}^{N} \alpha_j \sigma(\mathbf{y}_j^{\mathsf{T}} \mathbf{x} + \theta_j),$$

are dense in $C(I_n)$.

Definitions

Definition (Discriminatory functions)

We say σ is *discriminatory* if given a measure μ on I_n

$$\int_{I_n} \sigma(\mathbf{y}^{\mathsf{T}}\mathbf{x} + \theta)\, d\mu(\mathbf{x}) = 0$$

for all $\mathbf{y} \in \mathbf{R}^n$ implies $\mu = 0$.

Definition (Sigmoidal functions)

We say that σ is sigmoidal if

$$\begin{cases} \sigma(t) \to 1 & \text{as } t \to +\infty, \\ \sigma(t) \to 0 & \text{as } t \to -\infty. \end{cases}$$

The proof will rely on the following more general theorem:

Theorem

Let σ be continuous and discriminatory. Then the set of finite sums of the form

$$\sum_{j=1}^{N} \alpha_j \sigma(\mathbf{y}_j^{\mathsf{T}} \mathbf{x} + \theta_j),$$

is dense in $C(I_n)$.

Some Functional Analysis

The proof of the more general theorem relies on two important results from Functional Analysis; the Hahn-Banach Theorem and the Riesz Representation theorem.

Theorem

lol cats

A corrolary of this is what we use:

Theorem

lol kitty cats

The intuitive understanding you should gather from this is that if we have a linear subspace of a Banach space, we can find a nonzero linear functional which is zero on that linear subspace.

Now, the Riesz Representation theorem:

Theorem

lol kitty kitty cats

The intuitive understanding you should gather from this is that linear functionals can be represented by a “filter” of sorts.

We can now begin the proof of Theorem 1: First, let S be the set of continuous functions that we can exactly represent with our neural network. Clearly, S is a linear subspace of $C(I_N)$. So, by Hahn-Banach, there is a nonzero bounded linear functional that vanishes on the closure of S (the closure of S is the set of functions we can approximate with our neural network).

Now, this linear functional has the form

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

But $\sigma(y^T x + \theta)$ is in S for all y and θ. So,

$$\int \sigma(y^T x + \theta) d\mu(x) = 0$$

for all y and θ. But this means that μ is identically zero, because σ was discriminatory. That is, our linear functional is zero everywhere, which is a contradictiont c!

Next we need to show that sigmoidal functions are in fact discriminatory.

Lemma

Any bounded, measurable sigmoidal function is discriminatory. In particular,any continuous sigmoidal function is discriminatory.

By our previous results the main theorem now follows as a corollary to theorem 1 and lemma 2. That is, we have proven

Theorem

For σ a continuous sigmoidal function, finite sums of the form in Equation (1), *i.e.,*

$$\sum_{j=1}^{N} \alpha_j \sigma(\mathbf{y}_j^{\mathsf{T}} \mathbf{x} + \theta_j),$$

are dense in $C(I_n)$.