Approximation by Superpositions of a Sigmoidal Function

M. Ruby, C. Salinas

Department of Mathematics
Purdue University

Purdue University Machine Learning Seminar, 2019

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

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

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

The Main Result

The main result is the following theorem

Theorem

For σ a continuous sigmoidal function, finite sums of the form

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

are dense in $C(I_n)$.

The proof will rely on the following more general result

Theorem

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

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

is dense in $C(I_n)$.

With this theorem at our disposal we need only show that sigmoidal functions are discriminatory; that is, we will show that

Lemma

Any continuous sigmoidal function is discriminatory.

To prove Theorem 1 we will need a little bit of Functional Analysis. The following result is a corollary of the Hahn–Banach Theorem.

Corollary (cf. Conway, p. 79)

If $\mathcal{X}$ is a normed space, $\mathcal{M} \subset \mathcal{X}$ a subspace, $x_0 \in \mathcal{X} \setminus \mathcal{M}$, and $d = \mathrm{dist}(x_0, M)$, then there exists an element f of $\mathcal{X}^*$ such that $f(x_0) = 1$, $f(x) = 0$ for all x in $\mathcal{M}$, and $\|f\| = d^{-1}$.

Intuitively, this corollary says that given a subspace of a Banach space, we can find a nonzero linear functional which is zero on that subspace.

Now, the Riesz Representation Theorem:

Theorem (Riesz Representation Theorem)

If ϕ is a bounded linear functional on a Hilbert space $\mathscr{H}$, then there exists an element g of $\mathscr{H}$ such that for every f in $\mathscr{H}$,

$$\phi(f) = \langle f, g \rangle.$$

Definitions

Let's define a couple of these terms.

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 $\mu = 0$.

And of course.

Definition (Sigmoidal functions)

We say that σ is sigmoidal if

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

Proof of Theorem 1, Part 1

Let $S = \left\{ \sum_{j=1}^{N} \alpha_j \sigma(\mathbf{y}_j^{\mathsf{T}} \mathbf{x} + \theta_j) \right\} \subset C(I_n)$. Clearly S is a subspace of $C(I_n)$. We will show that the closure R of $\bar{S}$ is $C(I_n)$.
By the aforementioned corollary to the Hahn–Banach Theorem, there exist a functional L which is zero on S but nonzero on $C(I_n) \setminus S$.
Moreover, by the Riesz Representation Theorem, L is of the form

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

for some $\mu \in \mathcal{M}(I_n)$, for all $h \in C(I_n)$.

Proof of Theorem 1, Part 2

Since $\sigma(\mathbf{y}^{\mathsf{T}}\mathbf{x} + \theta)$ is in R for all $\mathbf{y}, \theta$,

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

for all $\mathbf{y}, \theta$.
However, since σ was assumed to be discriminatory, $\mu = 0$ contradicting the assumption that L is nonzero. Therefore, $R = C(I_n)$.

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 the sigmoidal functions that we care about are in fact discriminatory. That is,

Lemma

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

First we note that

$$\sigma(\lambda(\mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta)+\phi)\begin{cases}\to 1 & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta>0 \text{ as } \lambda\to+\infty,\\ \to 0 & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta<0 \text{ as } \lambda\to+\infty,\\ =\sigma(\phi) & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta=0 \text{ for all } \lambda.\end{cases}$$

So $\sigma_\lambda(\mathbf{x}) := \sigma(\lambda(\mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta)+\phi)$ converges pointwise boundedly to the function

$$\gamma(\mathbf{x})=\begin{cases}1 & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta>0,\\ 0 & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta<0,\\ \sigma(\phi) & \text{for } \mathbf{y}^{\mathsf{T}}\mathbf{x}+\theta=0,\end{cases}$$

as $\lambda\to\infty$.

Now, let $\Pi_{\mathbf{y},\theta} = \{\mathbf{x} \colon \mathbf{y}^{\mathsf{T}}\mathbf{x} + \theta = 0\}$ and let $H_{\mathbf{y},\theta} = \{\mathbf{x} \colon \mathbf{y}^{\mathsf{T}}\mathbf{x} + \theta > 0\}$. Then, by the Lebesgue Bounded Convergence Theorem,

$$\begin{aligned} 0 &= \int_{I_n} \sigma_\lambda(\mathbf{x})\, d\mu(\mathbf{x}) \\ &= \int_{I_n} \gamma(\mathbf{x})\, d\mu(\mathbf{x}) \\ &= \sigma(\phi)\mu(\Pi_{\mathbf{y},\theta}) + \mu(H_{\mathbf{y},\theta}) \end{aligned}$$

for all $\phi, \theta, \mathbf{y}$.

Proof of Lemma 1, Part 3

The last thing we need to show is that if $\mu(H_{\mathbf{y},\theta}) = 0$ for all $\mathbf{y}, \theta$ implies $\mu = 0$.
To that end, fix $\mathbf{y}$ and define the linear functional F as follows

$$F(h) = \int_{I_n} h(\mathbf{y}^T\mathbf{x})\, d\mu(\mathbf{x}).$$

Note that F is a bounded functional on $L^\infty(\mathbf{R})$ since μ is a finite signed measure.
Let $h = \chi_{[\theta,\infty)}$. Then

$$F(h) = \int_{I_n} \chi_{[\theta,\infty)}(\mathbf{y}^{\mathsf{T}}\mathbf{x}) = \mu(\Pi_{\mathbf{y},-\theta}) + \mu(H_{\mathbf{y},-\theta}) = 0.$$

Since the simple functions are dense in $L^\infty(\mathbf{R})$, $F = 0$.

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 (??), i.e.,

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

are dense in $C(I_n)$.