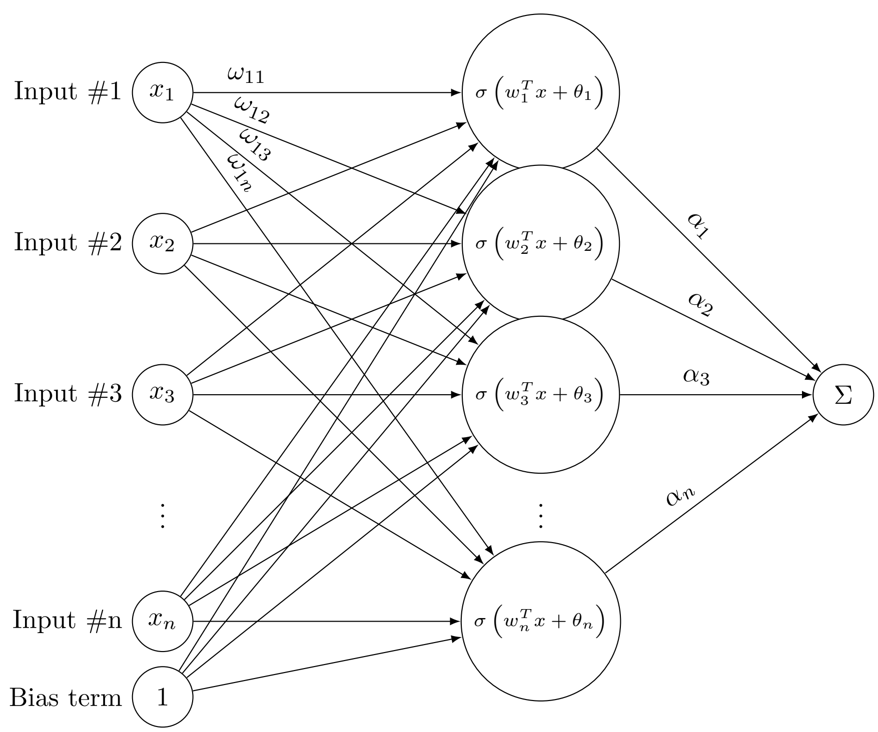
Neural network is universal approximator

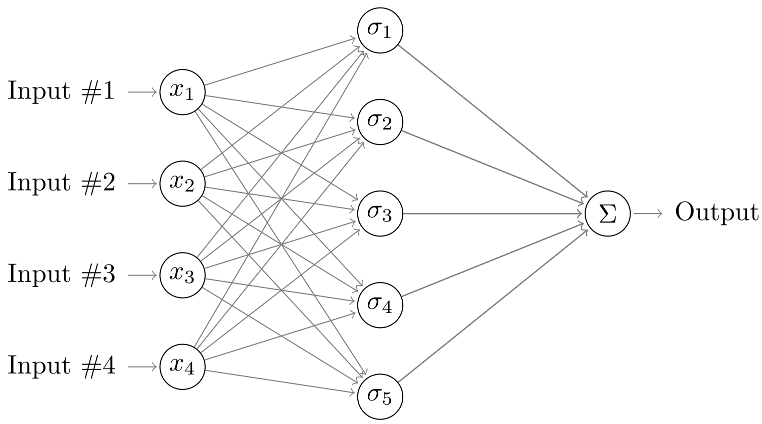
April 9, 2018

Plan:

- wstp o sigmoidzie wraz z grafik
- definicje i twierdzenia (hahn-banach, riesz)
- twierdzenie i dowd o gstoci kombinacji liniowej sigmoid
- dowd graficzny
- cytowania

0.0.1





Neural networks with sigmoidal activation functions can approximate to arbitrary accuracy any functional continuous mapping from one finite-dimensional space to another, provided the number N of hidden units is sufficiently large.

wiki: "A sigmoid function is a mathematical function having a characteristic "S"-shaped curve or sigmoid curve. Often, sigmoid function refers to the special case of the logistic function defined by the formula"

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

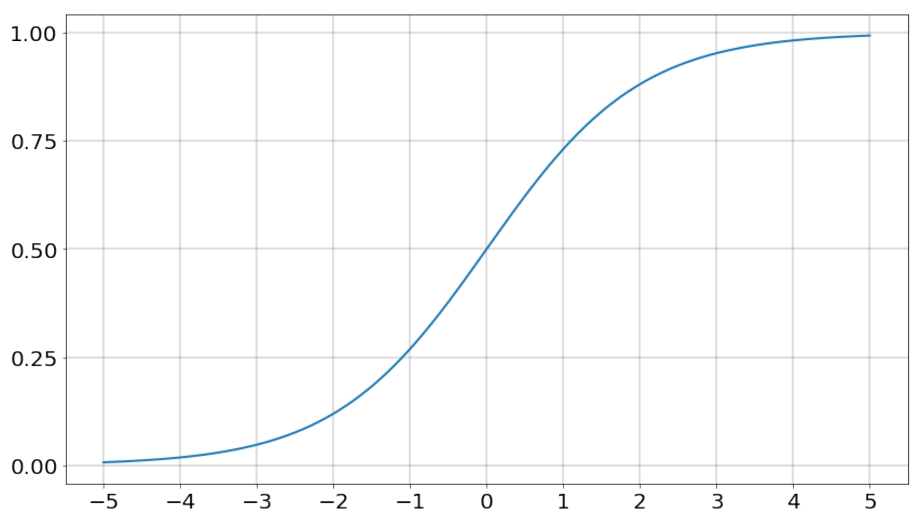


Figure 1: Sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$

Let I_n denote the n-dimensional unit cube, $[0,1]^n$. The space of continous functions on I_n is denoted by $C(I_n)$ and we use $||f||$ to denote the supremum norm of an $f \in C(I_n)$. The space of finite, signed regular Borel measures on I_n is denoted by $M(I_n)$.

Definition 0.1. We say that σ is sigmoidal if

$$\sigma(x) \to \begin{cases} 1 & \text{as} \quad x \to +\infty \\ 0 & \text{as} \quad x \to -\infty \end{cases}$$

Definition 0.2. We say that σ is discriminatory if for a measure $\mu \in M(I_n)$

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

for all $y \in \mathbf{R}$ and $\theta \in \mathbf{R}$ implies that $\mu = 0$.

Hahn-Banach theorem shows how to extend linear functionals from subspaces to whole spaces. Moreover, we can do it in a way that respects the boundedness properties of the given functional. The most general formulation of the theorem requires a preparation

Definition 0.3. A sublinear functional is a function $f : V \to \mathbf{R}$ on a vector space V which satisfies subadditivity (1) and positive homogenity conditions (2)

$$f(x+y) \le f(x) + f(y) \qquad \forall x, y \in V \tag{1}$$
$$f(\alpha x) = \alpha f(x) \qquad \forall \alpha \ge 0, x \in V \tag{2}$$

Theorem 0.1 (Hahn-Banach theorem for real vector spaces)**.** *If $p : V \to \mathbf{R}$ is a sublinear function, and $\psi : U \to \mathbf{R}$ is a linear functional on a linear subspace $U \subset V$, and satisfying $\psi(x) \le p(x)\ \forall x \in U$. Then there exists a linear extension $\Psi : V \to \mathbf{R}$ of ψ to the whole space V, such that*

- $\Psi(x) = \psi(x)\ \forall x \in U$
- $\Psi(x) \le p(x)\ \forall x \in V$

Rudin 1991, Th 3.2

Theorem 0.2 (Riesz representation theorem)**.** *Let H be a Hilber space over $\mathbf{R}$, and T a bounded linear functional on H. If T is a bounded linear functional on a Hilbert space H then there exist some $g \in H$ such that for every $f \in H$ we have (http://www.math.jhu.edu/ lindblad/632/riesz.pdf)*

$$T(f) = \langle f, g \rangle \qquad \forall f \in H$$

Any bounded linear functional T on the space of compactly supported continuous functions on X is the same as integration against a measure μ. (http://mathworld.wolfram.com/RieszRepres

$$Tf = \int f d\mu$$

Theorem 0.3. *Let σ be any continous discriminatory function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma \left(w_j^T x + \theta_j \right)$$

are dense in $C(I_n)$. In other words, given any $f \in C(I_n)$ and $\epsilon > 0$, there is a sum, $G(x)$, of the above form, for whic

$$|G(x) - f(x)| < \epsilon \qquad \forall x \in I_n$$

Proof. Let $S \subset C(I_n)$ be the set of functions of the form $G(x)$. Clearly S is a linear subspace of $C(I_n)$. We claim that the closure of S is all of $C(I_n)$.

Assume that closure of S is not all of $C(I_n)$. Then the closure of S, say R, is a closed proper subspace of $C(I_n)$. By the Hahn-Banach theorem, there is a bounded linear functional on $C(I_n)$, call it L, with the property that $L \neq 0$ but $L(R) = L(S) = 0$.

By the Riesz Representation Theorem, this bounded linear functional, L, is of the form

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

for some $\mu \in M(I_n)$, for all $h \in C(I_n)$. In particular, since $\sigma(y^T x + \theta)$ is in R for all y and θ, we must have that

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

for all y and θ.

However, we assumed that σ was discriminatory so that this condition implies that $\mu = 0$ contradicting our assumpition. Hence, the subspace S must be dense in $C(I_n)$.

This demonstrates that sums of the form

$$G(x) \sum_{j=1}^{N} \alpha_j \sigma \left(y_j^T x + \theta_j \right)$$

are dense in $C(I_n)$ providing that σ is continous and discriminatory.

$\square$

0.1 visual proof