# THE UNIVERSAL APPROXIMATION THEOREM

A first result is due to Cybenko (1989).

We give a proof for the one-dimensional case in steps.

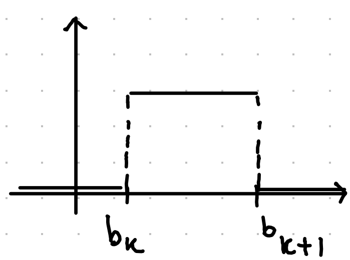
<u>Def</u>: A step function is a function $s:[a,b] \to \mathbb{R}$ of the form

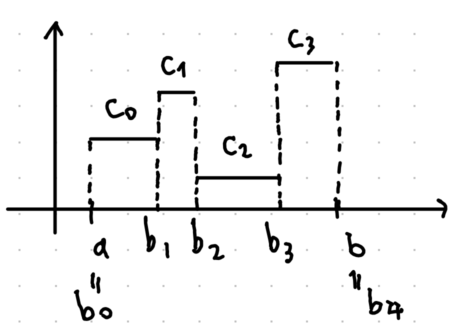
$$s(x) = \sum_{k=0}^{K-1} c_k \mathbb{1}_{[b_k, b_{k+1})}(x), \quad c_k \in \mathbb{R}$$

(include $b_k$)

where

$\mathbb{1}_{[b_k, b_{k+1})}(x)$ is the indicator function




example of a step function

<u>Theorem</u>: Let $f:[a,b] \longrightarrow \mathbb{R}$ be a continuous function. Then for every $\varepsilon > 0$ there exists a step function $s:[a,b] \to \mathbb{R}$ such that

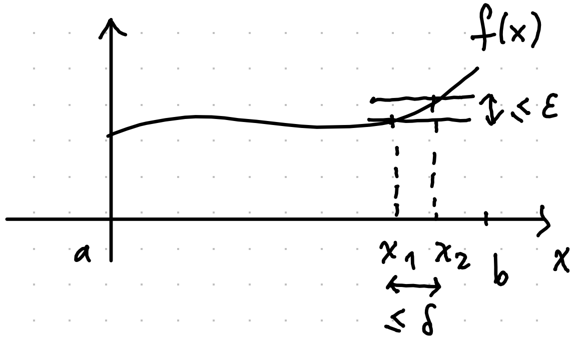
$$\sup_{x \in [a,b]} |s(x) - f(x)| \leq \varepsilon$$

<u>Proof</u>: By the Heine-Cantor Theorem, $f$ is uniformly continuous. This means that, given $\varepsilon > 0$, there exists a $\delta > 0$ (depending on $\varepsilon$) such that for every $x_1, x_2 \in [a,b]$ with $|x_1 - x_2| \leq \delta$ it holds

$$|f(x_1) - f(x_2)| \leq \varepsilon$$

(This is a classical result in Calculus. If you don't believe it, take $f$ to be a Lipschitz function:

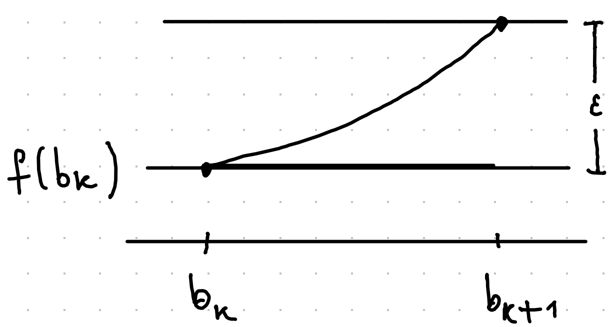
$$|f(x_1) - f(x_2)| \leq L\,|x_1 - x_2|$$ )




We partition $[a,b]$ using intervals of length $\delta$ (or less):

$$a = b_0 < b_1 < \dots < b_K = b$$

$$|b_{k+1} - b_k| \leq \delta$$




$$c_k = f(b_k)$$

and define the step function

$$s(x) = \sum_{k=0}^{K-1} c_k \, \mathbb{1}_{[b_k, b_{k+1})}(x)$$

Let us show that:

$$\sup_{x \in [a,b]} |f(x) - s(x)| \leq \varepsilon$$

Fix $x \in [a,b]$. It belongs to an interval,

$b_k \leq x < b_{k+1}$

Hence $s(x) = f(b_k)$.

Since $|x - b_k| \leq \delta$, we have that (by uniform continuity)

$$|f(x) - s(x)| = |f(x) - f(b_k)| \leq \varepsilon.$$

Since $x \in [a,b]$ is arbitrary, we can take the supremum and conclude the proof.

$\square$

See Python notebook.

<u>Remark</u>: If $\sup_{x \in [a,b]} |s(x) - f(x)| \leq \varepsilon$, then also integral errors can be made small. For example:

$$\int_a^b |s(x) - f(x)| \, dx \leq (b-a) \cdot \varepsilon$$

But also:

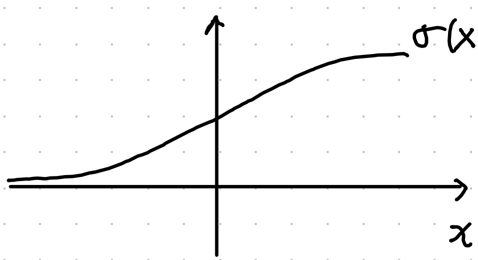
$$\int_a^b |s(x) - f(x)|^p \, dx \leq (b-a) \, \varepsilon^p$$

Hence, uniform distance small is a very strong approximation.

SIGMOID

The sigmoid function is

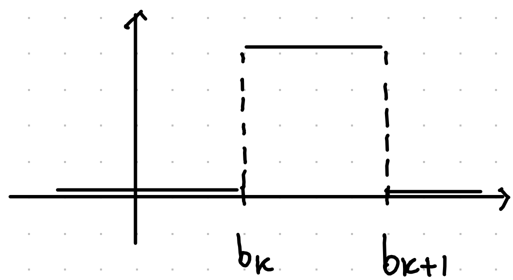
$$\sigma(x) = \frac{1}{1 + e^{-t}}$$



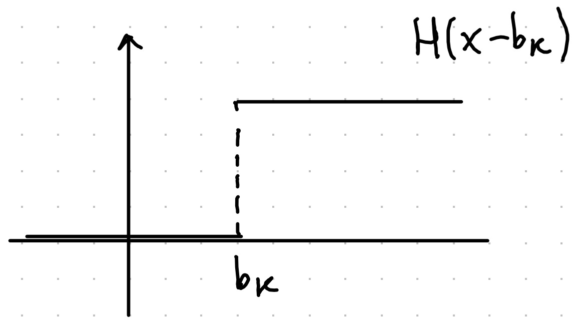
We want to approximate $f$ using sigmoids.

To do so, we observe that

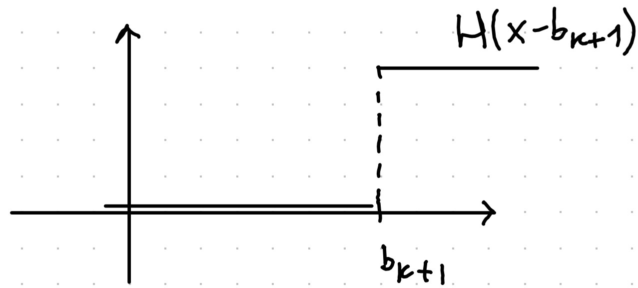
$$\mathbb{1}_{[b_k, b_{k+1})}(x) = H(x - b_k) - H(x - b_{k+1})$$




$=$




$-$




$H(x)$ is the Heaviside function.

Using these building blocks, we have to modify the constants $c_k$:

$$s(x) = \sum_{k=0}^{K-1} c_k \mathbb{1}_{[b_k, b_{k+1})}(x) = \sum_{k=0}^{K-1} c_k \left( H(x-b_k) - H(x-b_{k+1}) \right)$$

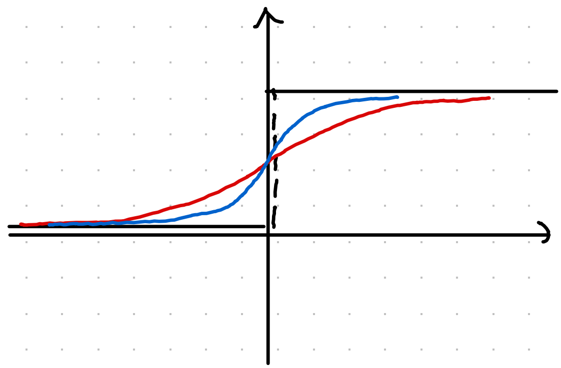
$$= \sum_{k=0}^{K-1} c_k H(x-b_k) - \sum_{k=0}^{K-1} c_k H(x-b_{k+1}) =$$

$$= \sum_{k=0}^{K-1} c_k H(x-b_k) - \sum_{k=1}^{K} c_{k-1} H(x-b_k) =$$

$$= c_0 H(x-b_0) + \sum_{k=1}^{K-1} (c_k - c_{k-1}) H(x-b_k) - c_K H(x-b_K)$$

$$= \sum_{k=0}^{K} d_k H(x-b_k)$$

We can approximate a Heaviside function with a sigmoid



To do so we consider $\sigma(wx)$.

Note that

$$\sigma(wx) \to H(x) \quad \text{as} \quad w \to +\infty$$

Hence we can use

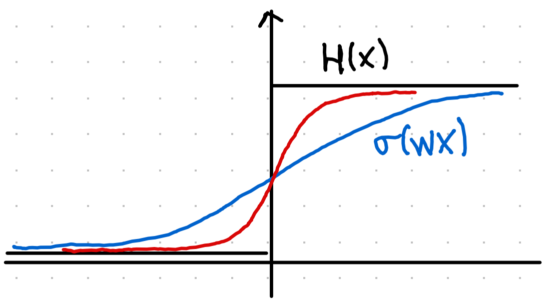
$$\sum_{k=0}^{K} d_k \, \sigma(w_k(x-b_k))$$

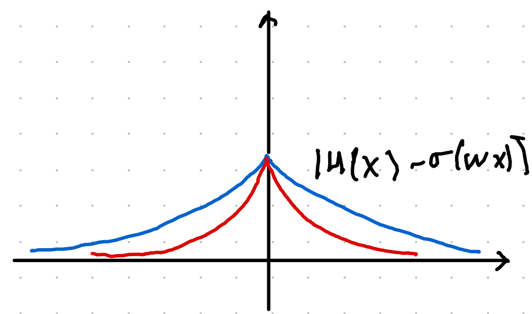
can approximate the Heaviside functions translated!

Unfortunately, we cannot use uniform distance with this argument, since $|H(x)-\sigma(wx)| = \frac{1}{2}$.

We can derive with almost no effort a bound on integral errors.

In fact, consider the error:




$$\int_{\mathbb{R}} |H(x)-\sigma(wx)|\,dx = \int_{-\infty}^{0} \sigma(wx)\,dx + \int_{0}^{+\infty} (1-\sigma(wx))\,dx$$

$$= \int_{-\infty}^{0} \frac{1}{1+e^{-wx}}\,dx + \int_{0}^{+\infty} 1 - \frac{1}{1+e^{-wx}}\,dx =$$

$$= \int_{-\infty}^{0} \frac{1}{1+e^{-wx}}\,dx + \int_{0}^{+\infty} \frac{e^{-wx}}{1+e^{-wx}}\,dx$$

$$= \int_{-\infty}^{0} \frac{1}{1+e^{-wx}}\,dx + \int_{0}^{+\infty} \frac{1}{e^{wx}+1}\,dx =$$

$$= 2\int_{0}^{+\infty} \frac{1}{1+e^{wx}}\,dx = \{ z = wx,\ dz = w\,dx \} =$$

$$= \frac{2}{|w|} \int_0^{+\infty} \frac{1}{1+e^z}\, dz = C \frac{1}{|w|} \longrightarrow 0$$

as $|w| \to +\infty$

Hence the quantities

$$\int_{\mathbb{R}} | d_k H(x-b_k) - d_k \sigma(w_k(x-b_k)) | \, dx$$

can be made as small as we want.

We can thus estimate

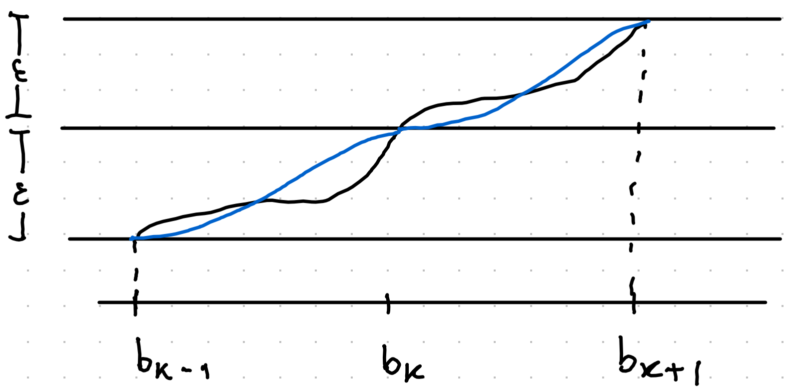
$$\int_a^b | f(x) - \sum_{k=0}^{K} d_k \sigma(w_k(x-b_k)) | \, dx \leq$$

$$\leq \underbrace{\int_a^b | f(x) - \sum_{k=0}^{K} d_k H(x-b_k) | \, dx}_{\text{can be made small}} +$$

$$+ \underbrace{\sum_{k=0}^{K} \int_{\mathbb{R}} | d_k H(x-b_k) - d_k \sigma(w_k(x-b_k)) | \, dx}_{\text{can be made small}}$$

$$\leq \varepsilon .$$

In fact, we could also go through the proof of the approximation via step functions and substitute directly in the proof step functions with sigmoids.




However this is technical. An "easy" proof requires some knowledge of functional analysis.

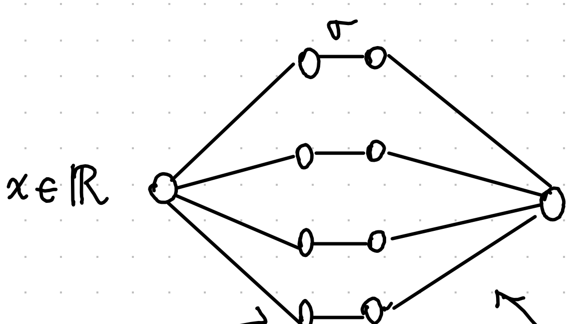
This

$$\sum_{k=0}^{K-1} d_k \, \sigma(w_k(x-b_k))$$

is nothing but a (shallow) neural network!

Let us reorganize the terms to write it as:

$$\sum_{k=0}^{K-1} \left( \sigma(x w^1_{1K} + b^1_K) w^2_{K1} + b^2 \right)$$




Linear layer
$W^1 \in \mathbb{R}^{1 \times K}$, $b^1 \in \mathbb{R}^{1 \times K}$

Linea layer
$W^2 \in \mathbb{R}^{K \times 1}$, $b^2 \in \mathbb{R}$

$xW + b$

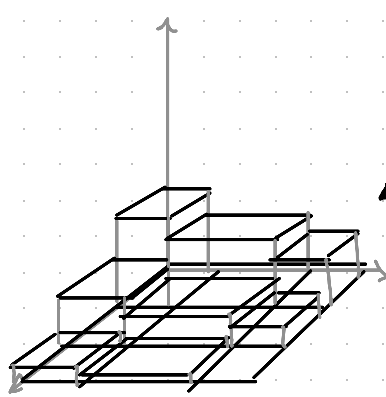
2D CASE

We will not go through the matematical ideas in this case, since the principles are the same. It just becomes more technical.

However we can give an idea.

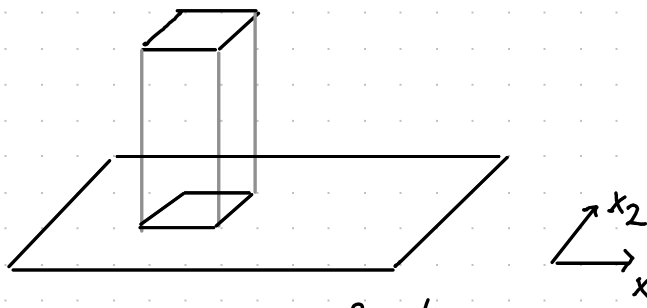
First of all, continuous functions can be approximated by step functions, i.e., fuctions that are constant on rectangles.

$$s(x_1, x_2) = \sum_{k=0}^{K-1} c_k \mathbb{1}_{\underbrace{R_k}_{\text{rectangle}}}(x_1, x_2)$$



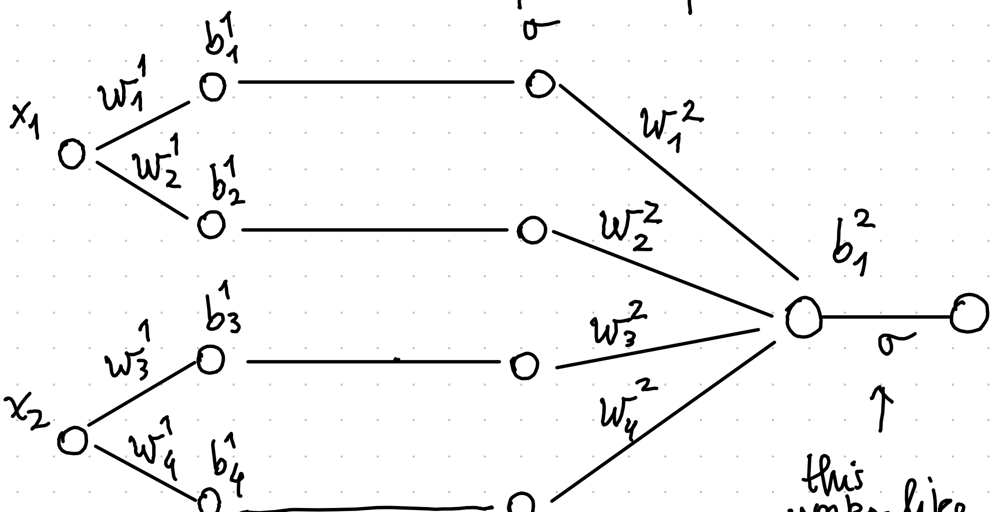
example of a step function in 2D

As we did in 1D, we approximate a single step using sigmoids:




a single step

We can approximate this with a neural network of this form:



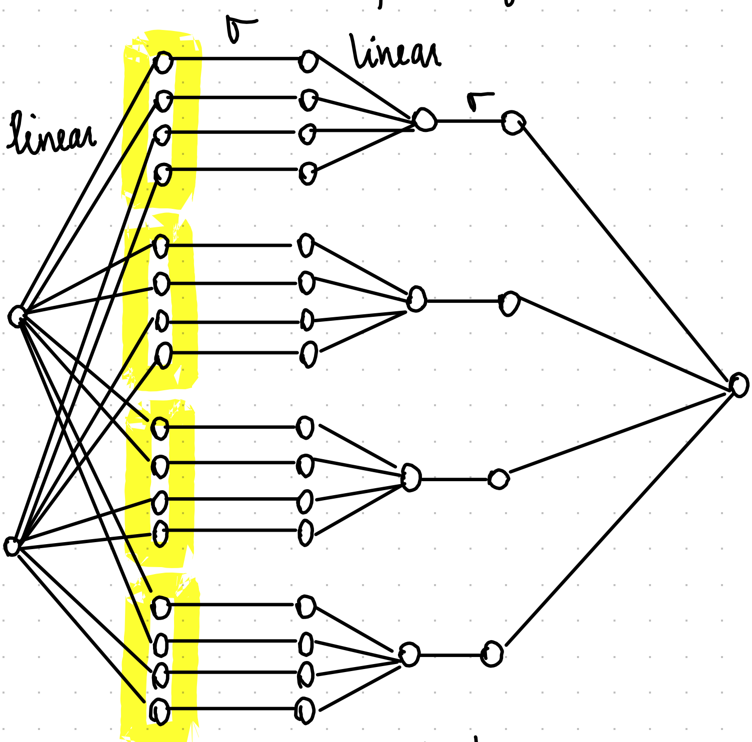

this works like a threshold check

See Python notebook.

This neural network has the following structure:

$$\sigma\Big(w_1^2\,\sigma(w_1^1 x_1^1 + b_1^1) + w_2^2\,\sigma(w_2^1 x_1 + b_2^1) + w_3^2\,\sigma(w_3^1 x_2 + b_3^1) + w_4^2\,\sigma(w_4^1 x_2 + b_4^1) + b_1^2\Big)$$

If we want to put together more steps:




4 step functions created