Machine Learning - DM3
About the Interest of Depth for Approximation by Neural Networks

due: April 9th, 2021

We focus in this exercise on ReLU neural networks defined on the input set $[0,1]$. We denote by $r : \mathbb{R} \to \mathbb{R}$ the ReLU function defined by $r(x) = \max(x, 0)$. We further denote by

$$\|h\|_\infty = \sup_{0 \le x \le 1} \big|h(x)\big|$$

the infinity norm of any function $h : [0,1] \to \mathbb{R}$.

General Approximation with Depth 2

Let $g : [0,1] \to \mathbb{R}$ be a function such that $g(0) = 0$ and g is L-Lipschitz: $\forall x, y \in [0,1], \big|g(x) - g(y)\big| \le L|x - y|$. A depth-2 ReLU network with linear output of width p is a function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \sum_{j=1}^{p} \alpha_j r\big(w_j x + b_j\big)$$

for some real parameters $w_1, \ldots, w_p$, $b_1, \ldots, b_p$, and $\alpha_1, \ldots, \alpha_p$.

1. Show that for every $\epsilon > 0$ there exists a depth-2 ReLU network f with linear output of width at most L/ϵ and parameters at most equal to $\max(2L, 1)$ such that $\|f - g\|_\infty \le \epsilon$.

The Sawteeth Function

We consider the function $s : [0,1] \to [0,1]$ defined by

$$s(x) = \begin{cases} 2x & \text{if } 0 \le x \le \frac{1}{2}\ , \\ 2 - 2x & \text{if } \frac{1}{2} \le x \le 1\ . \end{cases}$$

We further define for all $m \ge 1$ the function $s_m = \underbrace{s \circ \cdots \circ s}_{m \text{ times}}$.

2. Show that for all $x \in [0,1]$,

$$s(x) = 2\, r(x) - 4\, r\left(x - \frac{1}{2}\right)\ .$$

3. Plot s_1, s_2 and s_3 on the same figure.

4. Show that for all $m \geq 1$, all $k \in \{0, \ldots, 2^{m-1} - 1\}$ and all $t \in [0,1]$,

$$s_m\left(\frac{k+t}{2^{m-1}}\right) = s(t) \ .$$

5. For each $m \geq 1$, deduce from the preceeding questions a ReLU neural network computing s_m. Plot it. What is its depth? How many neurons does it have?

The Square Function

Let $g : [0,1] \to \mathbb{R}$ be the square function: $g(x) = x^2$. We show in this section that it can be approximated on $[0,1]$ by ReLU networks related to the sawteeth function.

For $m \geq 0$ let $g_m : [0,1] \to \mathbb{R}$ be such that $\forall k \in \{0, \ldots, 2^m\}$:

- $g_m\left(\frac{k}{2^m}\right) = g\left(\frac{k}{2^m}\right)$,
- g_m is linear on $\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right]$.

6. Plot g, g_0, g_1 and g_2 on the same figure.

7. Prove that for all $k \in \{0, \ldots, 2^m - 1\}$ and all $t \in [0,1]$,

$$g_m\left(\frac{k+t}{2^m}\right) - g\left(\frac{k+t}{2^m}\right) = \frac{t(1-t)}{4^m} \ .$$

8. Show that $\|g - g_m\|_\infty = \frac{1}{4^{m+1}}$ and for all $m \geq 1$,

$$g_m = g_{m-1} - \frac{1}{4^m}\ s_m = g_0 - \sum_{j=1}^{m} \frac{1}{4^j}\ s_j \ .$$

9. Deduce from the previous questions that for every $\epsilon > 0$, there exists a neural network f of depth $\lceil \log_4(1/\epsilon) \rceil$, and coefficients in $[-4, 2]$ such that $\|f - g\|_\infty \leq \epsilon$ on $[0,1]$. What is its width?

10. Compare this network with the one resulting from the approximation of Question 1, and comment.