TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Representing Functions with Shallow Circuits

The Classical Universality Theorems

Function Representations

Consider continuous functions $f : [0,1]^N \to \mathbb{R}$

$$\xrightarrow{f} \quad \mathbb{R}$$

Given the corner values, the interior can be filled.

$$f(x_1, \ldots, x_N) = E_{y_1,\ldots,y_N \sim \mathrm{Round}(x_1,\ldots,x_N)}\, f(y_1, \ldots, y_n)$$

Hence each of the 2^N corners has an independent value.

The Kolmogorov-Arnold representation theorem (1956)

For continuous $f : [0,1]^N \to \mathbb{R}$ there exists continuous "activation functions" $\sigma_i : \mathbb{R} \to \mathbb{R}$ and continuous $w_{i,j} : \mathbb{R} \to \mathbb{R}$ such that

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{2N+1} \sigma_i \left(\sum_{j=1}^{N} w_{i,j}(x_j) \right)$$

A Simpler, Similar Theorem

For (possibly discontinuous) $f : [0, 1]^N \to \mathbb{R}$ there exists (possibly discontinuous) $\sigma, w_i : \mathbb{R} \to \mathbb{R}$.

$$f(x_1, \ \ldots, \ x_N) = \sigma\left(\sum_i w_i(x_i)\right)$$

Proof: Select w_i to spread out the digits of its argument so that $\sum_i w_i(x_i)$ contains all the digits of all the x_i.

Cybenko's Universal Approximation Theorem (1989)

For continuous $f : [0,1]^N \to \mathbb{R}$ and $\varepsilon > 0$ there exists

$$
\begin{aligned}
F(x) &= \alpha^\top \sigma(Wx + \beta) \\
&= \sum_i \alpha_i \sigma \left(\sum_j W_{i,j}\, x_j + \beta_i \right)
\end{aligned}
$$

such that for all x in $[0,1]^N$ we have $|F(x) - f(x)| < \varepsilon$.

How Many Hidden Units?

Consider Boolean functions $f: \{0,1\}^N \to \{0,1\}$.

For Boolean functions we can simply list the inputs $x^0, \ldots, x^k$ where the function is true.

$$f(x) = \sum_k \mathbf{1}[x = x^k]$$

$$\mathbf{1}[x = x^k] \approx \sigma\left(\sum_i W_{k,i} x_i + b_k\right)$$

A simpler statement is that any Boolean function can be put in disjunctive normal form.

Representing Functions as IO Tables

These universality theorems implicitly treat functions as tables of intput-output pairs.

END