Teorema da Aproximação Universal (Generalização de Funahashi)

Teorema 1 (Leshno et al., 1993)**.** *Seja $K \subset \mathbb{R}^n$ compacto e seja $\sigma : \mathbb{R} \to \mathbb{R}$ uma função de ativação* ***não polinomial****. Então, o conjunto de todas as funções da forma*

$$F(x) = \sum_{i=1}^{N} a_i \, \sigma(w_i^\top x + b_i),$$

com $N \in \mathbb{N}$, $a_i \in \mathbb{R}$, $w_i \in \mathbb{R}^n$ e $b_i \in \mathbb{R}$, é denso em $C(K)$, o espaço das funções contínuas sobre K com a norma uniforme. Em outras palavras, para qualquer $f \in C(K)$ e $\varepsilon > 0$, existe uma rede neural de uma camada escondida tal que

$$\sup_{x \in K} |F(x) - f(x)| < \varepsilon.$$

Generalização em relação a Funahashi (1989):

- Funahashi exigia que a ativação fosse **sigmoidal, contínua, analítica e limitada**.
- Leshno et al. (1993) mostraram que basta que a função de ativação **não seja um polinômio**, abrangendo funções não limitadas como ReLU, senóide ou exponencial.
- Essa condição é ao mesmo tempo **necessária e suficiente** para a universalidade.

Intuição: Enquanto o teorema de Funahashi garante que sigmoides conseguem aproximar qualquer função contínua em um conjunto compacto, esta versão mostra que *qualquer ativação não polinomial* é suficiente — o que explica por que redes modernas com ReLU também são universalmente aproximadoras.