Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Redes Neurais: Fundamentos Matemáticos e Aplicações

Gabriel A. Martins[1] Murilo R. Cândido[2]
FCT UNESP, Presidente Prudente, SP

A crescente adoção de algoritmos de redes neurais em problemas de classificação e reconhecimento de padrões tem demonstrado a necessidade de fundamentar tais técnicas em bases teóricas sólidas e rigorosas. Matematicamente, uma rede neural pode ser descrita como a composição de funções cujo objetivo é transformar uma entrada $\mathbf{x} \in \mathbb{R}^{d_1}$ em uma saída $y \in \mathbb{R}^{d_{L+1}}$.

Sejam

$$\mathbf{f}^{\ell}_{\mathbf{W}^{\ell}} : \mathbb{R}^{d_\ell} \to \mathbb{R}^{d_{\ell+1}}, \quad \ell = 1, 2, \ldots, L, \tag{1}$$

as funções responsáveis pela transformação entre as camadas. Dessa forma, a rede com L camadas pode ser formalmente representada por:

$$\mathbf{R}_{\mathbf{W}}(\mathbf{x}) = \mathbf{f}^{L}_{\mathbf{W}^{L}}\Big(\mathbf{f}^{L-1}_{\mathbf{W}^{L-1}}\big(\cdots \mathbf{f}^{1}_{\mathbf{W}^{1}}(\mathbf{x})\big)\Big). \tag{2}$$

Em (2), cada função $\mathbf{f}^{\ell}_{\mathbf{W}^{\ell}}$ depende dos parâmetros $\mathbf{W}^{\ell}$ e, em geral, apresenta comportamento não linear.

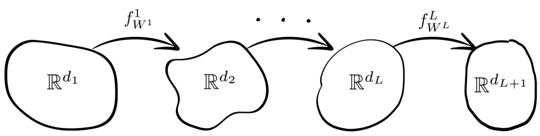

Figura 1: Representação esquemática da composição de funções em uma rede neural. Fonte: Elaborado pelos autores.

O embasamento deste trabalho apoia-se no Teorema da Aproximação Universal [**Cybenko1989**], que assegura a capacidade de uma rede neural em aproximar qualquer função contínua sob hipóteses razoáveis. A demonstração deste teorema utiliza resultados clássicos da análise funcional e teoria da medida, dentre os quais se destacam:

Teorema 1 (Hahn-Banach [3]). *Seja X um espaço vetorial real e $p : X \to \mathbb{R}$ um funcional sublinear. Se $Z \subset X$ for um subespaço e $f : Z \to \mathbb{R}$ um funcional linear que satisfaz*

$$f(x) \leq p(x) \quad \textit{para todo } x \in Z,$$

então existe uma extensão linear $\tilde{f} : X \to \mathbb{R}$ de f tal que

$$\tilde{f}(x) \leq p(x) \quad \textit{para todo } x \in X.$$

[1]gabriel.a.martins@unesp.br
[2]mr.candido@unesp.br

Teorema 2 (Convergência Dominada [1]). *Seja* $\{f_n\}$ *uma sequência de funções mensuráveis convergindo pontualmente para uma função* f *em um conjunto de medida finita, e suponha que exista uma função integrável* g *satisfazendo*

$$|f_n(x)| \leq g(x) \quad \textit{para todo } n \in \mathbb{N} \textit{ e quase todo } x.$$

Então,

$$\lim_{n\to\infty} \int f_n(x)\,dx = \int f(x)\,dx. \tag{3}$$

Teorema 3 (Representação de Riesz [2]). *Seja* H *um espaço de Hilbert. Para todo funcional linear contínuo* $f \in H^*$*, existe um único elemento* $y \in H$ *tal que*

$$f(x) = \langle x, y\rangle, \quad \textit{para todo } x \in H. \tag{4}$$

Esses teoremas formam a base para a demonstração do Teorema da Aproximação Universal, que fundamenta a capacidade de redes neurais em aproximar funções contínuas.

Com a base teórica estabelecida, demonstraremos o **Teorema da Aproximação Universal** e evidenciamos que sua compreensão aprofundada possibilita a construção de arquiteturas de redes neurais mais abrangentes do que os tradicionais perceptrons multicamadas,

$$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)\,(\mathbf{x}), \tag{5}$$

onde $\mathbf{W}_i$ representam matrizes de coeficientes e σ_i as funções de ativação, amplamente empregadas nos modelos atuais.

Adicionalmente, combinamos esse resultado com o teorema de Kolmogorov-Arnold, que assegura que qualquer função contínua de várias variáveis pode ser representada como uma soma de funções contínuas de uma única variável. Essa integração teórica fundamenta o desenvolvimento de arquiteturas de redes neurais mais gerais, capazes de superar as limitações inerentes às abordagens convencionais. Tais conceitos serão, então, aplicados na construção de um classificador de risco de evasão escolar para estudantes da UNESP.

A abordagem apresentada evidencia a relevância dos fundamentos matemáticos na modelagem de redes neurais, demonstrando como a comprrensão de teoremas clássicos pode ser útil na construção de modelos preditivos. Essa abordagem permite a elaboração de arquiteturas capazes de aproximar funções complexas, contribuindo para avanços nas técnicas de classificação em dados educacionais.

Agradecimentos

O presente trabalho é realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processos $2024/17594-5$ e $2023/06076-0$.

Referências

[1] R. G. Bartle. **The Elements of Integration and Lebesgue Measure**. New York: John Wiley & Sons, 2014. ISBN: 0-471-04222-6.

[2] E. Kreyszig. **Introductory Functional Analysis with Application**. John Wiley & Sons, 1991. ISBN: 0-471-50731-8.

[3] A. N. Kolmogorov e S. V. Fomin. **Elements of the Theory of Functions and Functional Analysis**. New York: Courier Corporation, 1957. ISBN: 0-486-40683-0.