

Teorema da Aproximação Universal

Caio Lins

16 de maio de 2021

Conteúdo

1	Introdução	3
---	------------	---

1 Introdução

Neste trabalho estudaremos resultados que tratam da tarefa de representar um conjunto de funções por meio de outro, podendo essa representação ser exata ou aproximada. Esse é um tema amplo, presente em diversas áreas da matemática, tanto pura quanto aplicada. Faremos esse estudo em alguns contextos diferentes, sendo o principal deles o da aproximação de funções contínuas por meio de redes neurais.

Começaremos pelo teorema da aproximação de Weierstrass, que estabelece a viabilidade de aproximar arbitrariamente bem funções contínuas em um intervalo da reta por meio de polinômios. Inicialmente provado por Weierstrass em 1885, novas demonstrações surgiram com o tempo. A que apresentaremos — retirada de [You06] — ainda fornecerá uma forma de, dada uma função contínua, obter os polinômios que a aproximam. Essa é uma peculiaridade desse teorema, pois as provas seguintes não são construtivas.

Escolhemos iniciar com esse resultado por sua demonstração ser elementar, não exigindo mais que um curso de análise na reta. Além disso, ele trata de aproximações em um contexto mais simples, com apenas uma variável e no qual as funções aproximadoras, os polinômios, são lineares com relação aos seus coeficientes, os parâmetros da aproximação. Com isso, esperamos acostumar o leitor ao tipo de questão que nos é de interesse.

Em seguida, moveremos nossa atenção para um dos problemas apresentados por David Hilbert no Congresso Internacional de Matemáticos de 1900, mais especificamente, o 13°. Hilbert postulou (utilizando a linguagem matemática de sua época) que existem funções contínuas de \mathbb{I}^3 em \mathbb{R} , onde $\mathbb{I} = [0, 1]$, que não podem ser expressas por meio da composição e adição de funções de \mathbb{R}^2 em \mathbb{R} . Décadas após ser postulada, essa conjectura eventualmente foi demonstrada *falsa*. A prova foi dada por Vladimir Igorevich Arnol'd, 14 anos após a morte de Hilbert. Ele e seu orientador de Doutorado, Andrej Nikolajewitsch Kolmogorov, provaram que, na verdade, toda função contínua $f : \mathbb{I}^n \rightarrow \mathbb{R}$ pode ser expressa como composições e adições de funções contínuas de $\mathbb{R} \rightarrow \mathbb{R}$. A formulação exata que apresentaremos — obtida de [Mor21] — é ainda mais forte que essa.

Naturalmente, a diferença mais óbvia desse resultado para o anterior é que agora nós temos representações *exatas* para nossas funções de interesse, não apenas aproximações. Além disso, esse é um contexto que envolve múltiplas variáveis e no qual as funções aproximadoras não são lineares em seus parâmetros. Portanto, é de se esperar que a demonstração apresentada seja mais complexa. De fato, apesar de não ser extremamente complicada, ela envolve conceitos e resultados sobre espaços métricos aos quais julgamos apropriado dedicar um apêndice.

O último e principal resultado sobre aproximações que apresentaremos é o teorema da aproximação universal, como formulado em [Cyb89]:

Teorema da aproximação universal. *Seja $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua discriminatória qualquer. Então, dada qualquer $f : \mathbb{I}^n \rightarrow \mathbb{R}$, contínua, e $\varepsilon > 0$, existe uma soma, $G : \mathbb{I}^n \rightarrow \mathbb{R}$, da seguinte forma:*

$$G(x) = \sum_{j=1}^N \alpha_j \sigma(y_j^T x + \theta_j), \quad (1)$$

onde $\alpha_j, \theta_j \in \mathbb{R}$ e $y_j \in \mathbb{R}^n$, tal que

$$|G(x) - f(x)| < \varepsilon$$

para todo $x \in \mathbb{I}^n$.

Aqui y_j^T é o transposto do vetor y_j , de modo que $y_j^T x$ é o produto interno usual de \mathbb{R}^n . A definição de função discriminatória será dada posteriormente.

Antes de apresentar comentários sobre esse teorema, é prudente explicar sua relação com redes neurais. Como descrito em [Lip87], as redes neurais artificiais são algoritmos de computação que surgiram como uma tentativa de espelhar o funcionamento de redes neurais orgânicas, como o cérebro humano. Em essência, seu objetivo é “atingir bom desempenho por meio da densa interconexão de elementos computacionais simples.”

Na prática, existem várias formas de atingir esse objetivo. No tipo de rede ao qual daremos enfoque, o processo de computação é realizado por um conjunto de nós, organizados em camadas ordenadas, em que a camada inicial é composta por n nós de **input** e a camada final, por m nós de **output**. Entre eles a rede pode possuir camadas intermediárias de tamanho variado. Dados n valores reais para os nós de **input**, a rede produz, nos nós de **output**, m valores, também reais. Logo, ela pode ser considerada uma função de \mathbb{R}^n em \mathbb{R}^m .

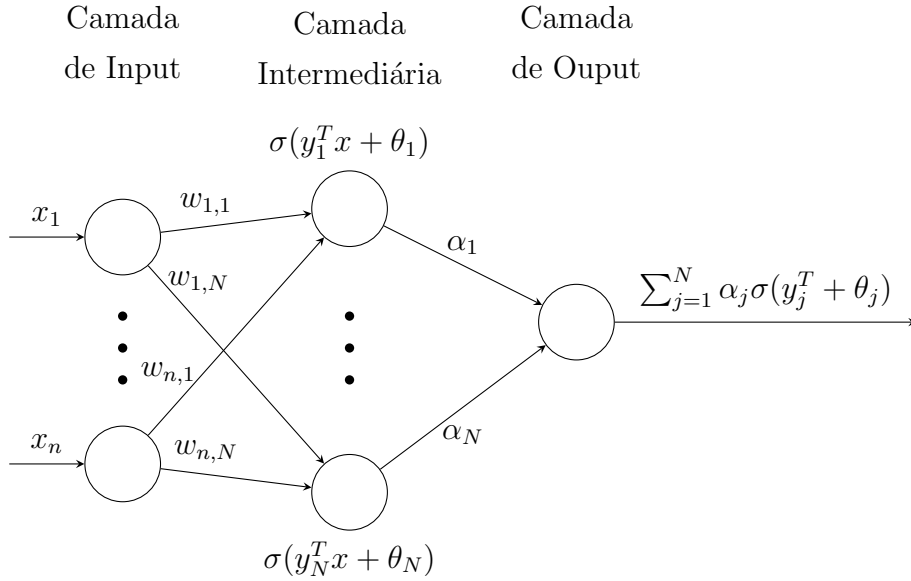


Figura 1: Rede neural com apenas uma camada intermediária. Aqui temos $x = [x_1 \ \cdots \ x_n]^T$ e $y_j^T x = [w_{1,j} \ \cdots \ w_{n,j}]$, o vetor dos pesos de cada nó intermediário. Na última camada ocorre apenas uma combinação linear.

A partir da segunda camada, cada nó está conectado a todos os nós da camada anterior por meio de uma aresta, a qual possui um peso. Esse nó computa uma combinação linear, utilizando como coeficientes os pesos das arestas, dos valores armazenados pelos nós da camada anterior, soma um fator de correção ao valor obtido e passa o resultado por uma função não-linear, obtendo assim um

valor próprio. O caso em que $m = 1$ e há apenas uma camada intermediária de N nós, ao qual nos atentaremos, está representado na figura 1, onde σ é a não-linearidade.

Repare que o output da rede é exatamente a expressão em (1). Ou seja, o teorema da aproximação universal, doravante conhecido como **TAU**, estabelece a possibilidade de aproximar arbitrariamente bem funções reais contínuas em \mathbb{I}^n por meio de redes neurais com uma camada intermediária. Essa é uma pergunta de extrema relevância, tanto teórica como prática pois, como apontado em [Lip87], redes neurais artificiais possuem diversas aplicações em campos voltados ao desenvolvimento de classificadores robustos, como a teoria de reconhecimento de fala e de imagens. Saber que é possível realizar aproximações arbitrariamente boas utilizando redes neurais artificiais dá mais segurança e incentivo à pesquisa nessas áreas.

A demonstração do **TAU** envolve resultados de teoria da medida e análise funcional, os quais são importantes por si mesmos e, por isso, são discutidos em seções próprias. Alguns conhecimentos necessários para entendê-los são apresentados como outro apêndice. De fato, a demonstração desse resultado é mais densa que a dos outros, apesar de que ele, ao contrário do problema de Hilbert, lida apenas com aproximações. Uma possível explicação para essa diferença será apresentada na seção correspondente.

Referências

- [Lip87] R. Lippmann. «An introduction to computing with neural nets». Em: *IEEE ASSP Magazine* 4.2 (1987), pp. 4–22. DOI: 10.1109/MASSP.1987.1165576.
- [Cyb89] George Cybenko. «Approximation by superpositions of a sigmoidal function». Em: *Math. Control. Signals Syst.* 2 (1989), pp. 303–314. DOI: 10.1007/BF02551274. URL: <https://doi.org/10.1007/BF02551274>.
- [You06] Matt Young. *The Stone-Weierstrass theorem*. Jan. de 2006.
- [Mor21] Sidney A. Morris. «Hilbert 13: are there any genuine continuous multivariate real-valued functions?» Em: *Bulletin (New Series) of the American Mathematical Society* 58.1 (jan. de 2021), pp. 107–118. URL: <https://doi.org/10.1090/bull/1698>.