O que realmente significa um modelo “aprender”?

A palavra aprendizado nasce da nossa experiência humana de observar, errar e ajustar.

2026-02-16

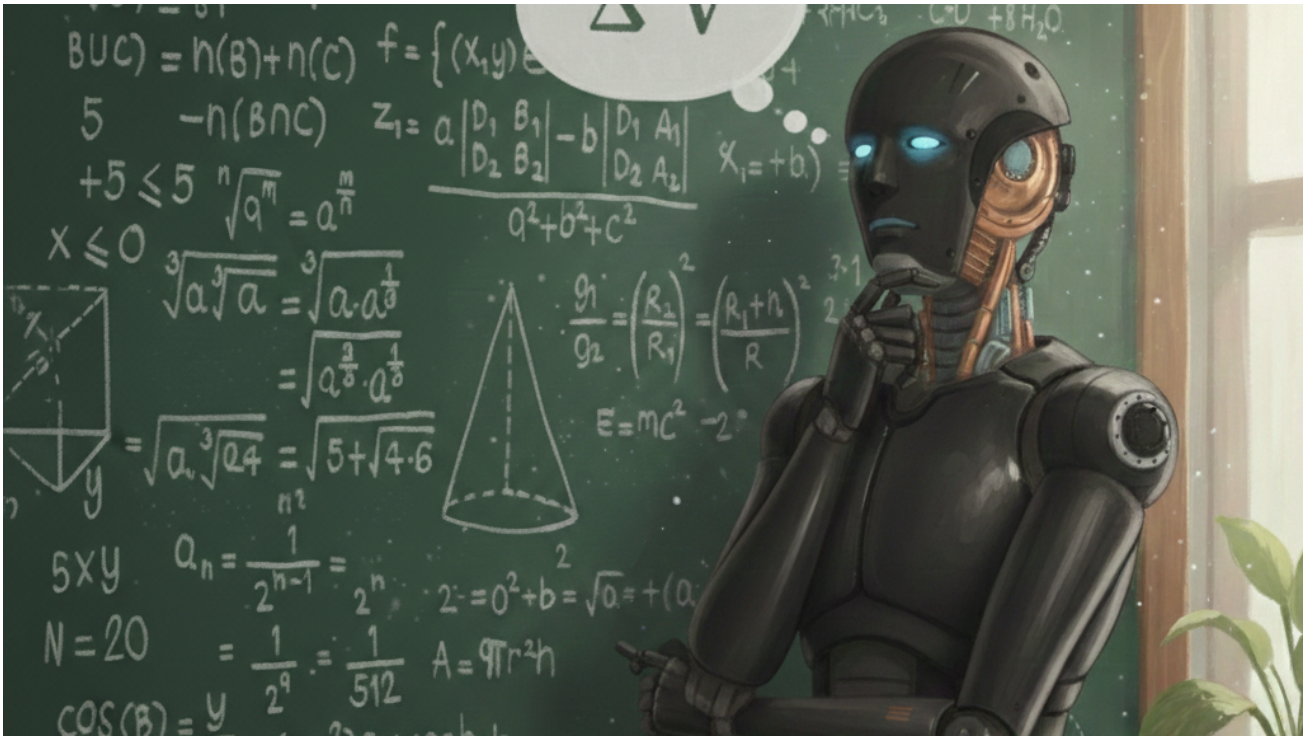


Figure 1: Capa do artigo

A palavra aprendizado nasce da nossa experiência humana de observar, errar e ajustar.

Na ciência da computação, ela passou a ter um significado técnico: melhorar desempenho a partir de dados.

Em 1959, Arthur Samuel descreveu aprendizado de máquina como a capacidade de um sistema melhorar seu desempenho em uma tarefa conforme recebe mais experiência [1].

Décadas depois, Tom Mitchell formalizou essa definição de forma objetiva: um programa aprende com Experiência (E) em relação a uma Tarefa (T) e uma medida de Desempenho (P) se seu desempenho melhora com E [2].

A Estrutura T, E, P A ideia pode ser organizada em três elementos:

- Tarefa (T): o que o sistema deve fazer
- Experiência (E): os dados que ele recebe
- Desempenho (P): a métrica que mede se está melhorando

Se o desempenho (P) melhora à medida que a experiência (E) aumenta em uma tarefa (T), houve aprendizado.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

FORMULAÇÃO MATEMÁTICA

O aprendizado pode ser descrito como a minimização de uma função de perda:

$$\theta^* = \arg\min_{\theta} L(\theta)$$

representa o conjunto de parâmetros do modelo L( ) representa a função de erro (ou função de perda)

Aprender significa encontrar o conjunto de parâmetros que torna o erro o menor possível.

O PAPEL DO GRADIENTE

O gradiente é representado por:

$$\nabla L(\theta)$$

Ele é o vetor das derivadas parciais da função de erro em relação aos parâmetros.

Em termos práticos, ele indica como pequenas variações em cada parâmetro alteram o valor do erro.

A atualização dos parâmetros ocorre segundo:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

onde:

Onde (eta) é a taxa de aprendizado, parâmetro que controla o tamanho do passo em cada atualização.

Valores maiores aceleram o processo, enquanto valores menores tornam os ajustes mais graduais e estáveis [3].

A descida do gradiente realiza esse processo de forma iterativa, atualizando os parâmetros na direção que reduz o erro de maneira consistente.

RETROPROPAGAÇÃO

Em redes neurais profundas, surge uma questão prática: como calcular L( ) para todos os parâmetros em todas as camadas de forma eficiente?

A retropropagação responde a isso. Ela utiliza a regra da cadeia do cálculo diferencial para computar o gradiente camada a camada, da saída para as partes internas do modelo [4].

Isso permite atualizar milhões ou bilhões de parâmetros de forma coordenada e viabiliza o treinamento moderno de redes profundas em larga escala [5].

GENERALIZAÇÃO

Um ponto central é a generalização: manter bom desempenho em dados que o modelo ainda não viu.

Por isso avaliamos modelos com conjuntos separados de validação e teste, justamente para medir essa capacidade [2].

CONCLUSÕES

No fim, o que chamamos de aprendizado é um processo de otimização orientado por dados.

Modelos ajustam funções matemáticas sob restrições específicas.

Compreender essa estrutura traz clareza sobre como avaliamos desempenho, limites e implicações de modelos de IA.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

REFERÊNCIAS

[1] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” IBM Journal of Research and Development, vol. 3, no. 3, pp. 210–229, 1959.

[2] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.

[3] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv:1609.04747, 2016.

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature, vol. 323, pp. 533– 536, 1986.

[5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

Postagem original no LinkedIn: O que realmente significa um modelo “aprender”?

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.