Redes neurais: uma breve história

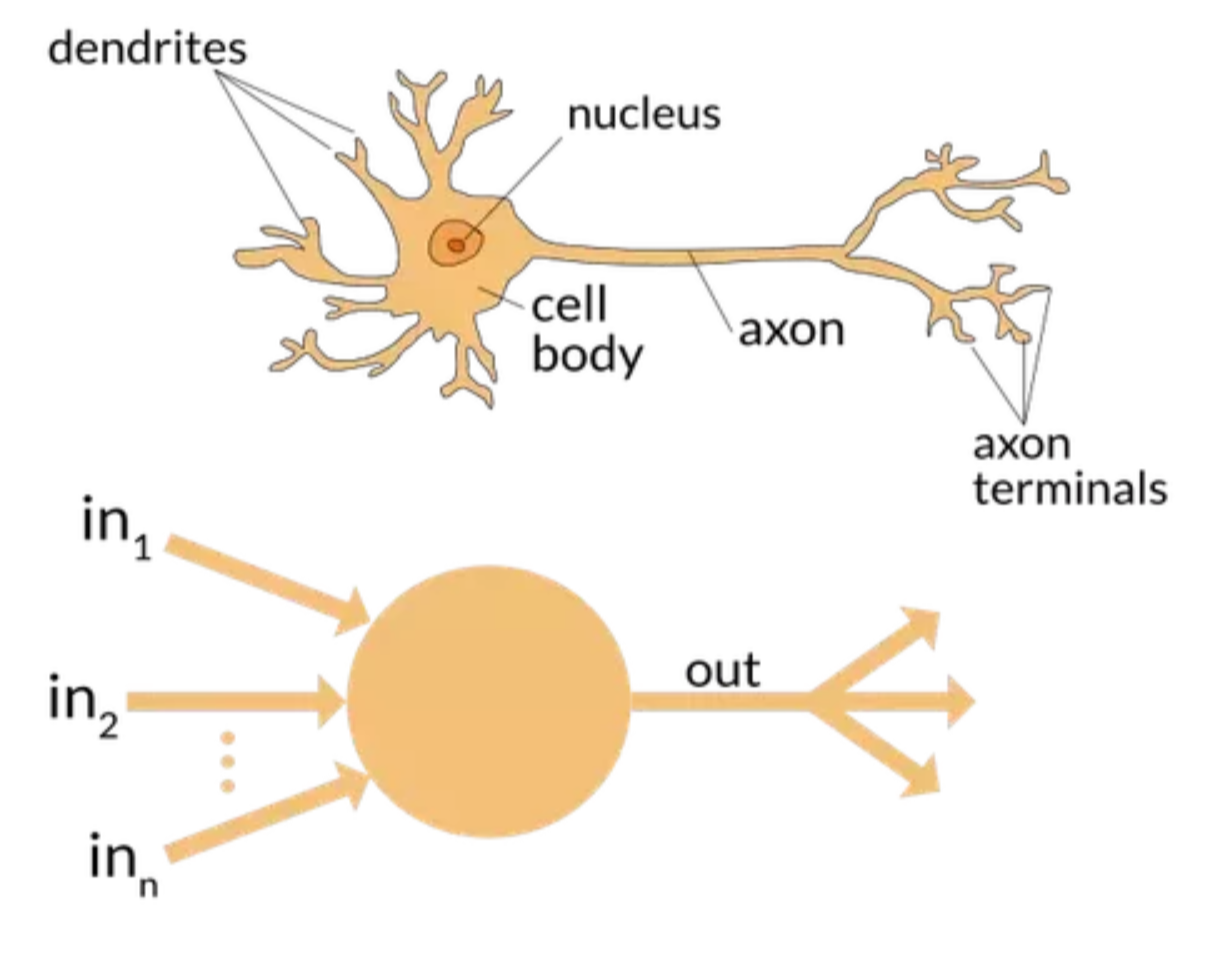
**Fonte**: tradução automática por Google Docs de “[Your deep learning journey](https://github.com/fastai/fastbook/blob/master/01_intro.ipynb)” do [livro Fastai](https://github.com/fastai/fastbook) (Jeremy Howard e Sylvain Gugger).

# Modelo matemático de um neurônio artificial

Em 1943, **Warren McCulloch**, um neurofisiologista, e **Walter Pitts**, um lógico, se uniram para desenvolver **um modelo matemático de um neurônio artificial**. Eles declararam que:

*Devido ao caráter "tudo ou nada" da atividade nervosa, os eventos neurais e as relações entre eles podem ser tratados por meio da lógica proposicional. Verifica-se que o comportamento de toda rede pode ser descrito nesses termos*. (Pitts e McCulloch; Um cálculo lógico das idéias imanentes à atividade nervosa)

Eles perceberam que **um modelo simplificado de um neurônio real poderia ser representado usando adição e limiares simples,** como mostrado aqui:



# O Perceptron: a capacidade de aprender

**Frank Rosenblatt** ainda mais( 1956) desenvolveu o neurônio artificial para lhe dar a **capacidade de aprender**. Mais importante ainda, ele trabalhou na construção do primeiro dispositivo que realmente usava esses princípios: o Mark I **Perceptron**.

Rosenblatt escreveu sobre este trabalho: "**estamos prestes a testemunhar o nascimento de tal máquina - uma máquina capaz de perceber, reconhecer e identificar seu entorno sem nenhum treinamento ou controle humano**". O perceptron foi construído e conseguiu reconhecer formas simples.

# Perceptrons: início do primeiro inverno da IA

Um professor do MIT chamado **Marvin Minsky** (que ficou atrás de Rosenblatt na mesma escola!) Junto com **Seymour Papert** escreveu um **livro, chamado "Perceptrons"**, sobre a invenção de Rosenblatt. Eles mostraram que uma única camada desses dispositivos não conseguiu aprender algumas funções matemáticas simples e críticas (como o XOR).

No mesmo livro, eles também mostraram que o uso de várias camadas dos dispositivos permitiria resolver essas limitações. Infelizmente, apenas a primeira dessas idéias foi amplamente reconhecida, como resultado da comunidade acadêmica global quase inteiramente desistindo de redes neurais pelas próximas duas décadas.

# Redes neurais

Talvez o trabalho mais crucial em redes neurais nos últimos 50 anos seja ovários volumes *Processamento Distribuído Paralelo* (PDP) de, lançado em 1986 pela MIT Press. Capítulo 1 estabelece uma esperança semelhante ao mostrado por

Rosenblatt:: *... as pessoas são mais espertos do que os computadores de hoje porque o cérebro emprega uma arquitetura computacional básico que é mais adequado para lidar com um aspecto central das informações naturais processamento de tarefas que as pessoas são tão bons às. ... introduziremos uma estrutura computacional para modelar processos cognitivos que parece ... mais próxima do que outras estruturas ao estilo de computação, como pode ser feito pelo cérebro.* (PDP, capítulo 1)

A premissa que o PDP está usando aqui é que **os programas de computador tradicionais funcionam de maneira muito diferente dos cérebros**, e pode ser por isso que os programas de computador (naquele momento) foram tão ruins em fazer coisas que os cérebros acham fáceis (como reconhecer objetos nas fotos). Os autores afirmam que a abordagem PDP é "mais próxima do que outras estruturas" de como o cérebro funciona e, portanto, pode ser mais capaz de lidar com esse tipo de tarefa.

De fato, a abordagem apresentada no PDP é muito semelhante à usada nas redes neurais atuais. O livro definia "Processamento Distribuído Paralelo" como exigindo:

1. Um conjunto dede *unidadesprocessamento*
2. Um *estado de ativação*
3. Uma *função de saída* para cada unidade
4. Um *padrão de conectividade* entre unidades
5. Uma *regra* de propagação para propagar padrões de atividades através da rede de conectividades
6. Uma *regra de ativação* para combinar as entradas que colidem com uma unidade com o estado atual dessa unidade para produzir um novo nível de ativação para a unidade.
7. Uma *regra de aprendizado* segundo a qual os padrões de conectividade são modificados pela experiência.
8. Um *ambiente* no qual o sistema deve operar.

Veremos neste livro que redes neurais lidam com cada um desses requisitos.

# Década de 1980: primeiras aplicações de NN

Na década de 1980, a maioria dos modelos foi construída com uma segunda camada de neurônios, evitando assim o problema identificado por Minsky (esse era seu "padrão de conectividade entre unidades", para usar a estrutura acima). E, de fato, as redes neurais foram amplamente usadas nas décadas de 80 e 90 para projetos reais e práticos. No entanto, novamente um equívoco das questões teóricas impediu o campo. Em teoria, adicionar apenas uma camada extra de neurônios era suficiente para permitir que qualquer modelo matemático se aproximasse dessas redes neurais, mas, na prática, essas redes eram muitas vezes grandes e lentas demais para serem úteis.

# Mais camadas!

Embora os pesquisadores tenham mostrado há 30 anos que, para obter um bom desempenho prático, você precisa usar ainda mais camadas de neurônios, é apenas na última década que isso tem sido mais amplamente apreciado. As redes neurais estão finalmente alcançando seu potencial, graças ao entendimento de usar mais camadas e à capacidade aprimorada de fazê-lo, graças a melhorias no hardware do computador, aumentos na disponibilidade de dados e ajustes algorítmicos que permitem que as redes neurais sejam treinadas mais rapidamente e mais facilmente.

Agora temos o que Rosenblatt havia prometido: "uma máquina capaz de perceber, reconhecer e identificar seu entorno sem nenhum treinamento ou controle humano".