

## Funções de Ativação em Redes Neurais

As **funções de ativação** são componentes fundamentais das redes neurais artificiais. Elas definem como o sinal de entrada, após a combinação linear dos pesos e do bias, será transformado antes de passar para a próxima camada. Em outras palavras, a função de ativação introduz **não-linearidade**, permitindo que a rede aprenda relações complexas entre as variáveis de entrada.

### 1. Por que precisamos de funções de ativação?

Se utilizássemos apenas combinações lineares, uma rede neural — mesmo com várias camadas — não teria maior capacidade de representação do que um único neurônio. A inclusão de funções não lineares possibilita que a rede resolva problemas que não são linearmente separáveis, como é o caso do famoso problema do **XOR**.

### 2. Principais funções de ativação

- **Degrau:** é a mais simples, produz apenas duas saídas possíveis (0 ou 1). Foi usada no Perceptron de Rosenblatt, mas não permite aprendizado via retropropagação porque não é diferenciável. *Se não compreendeu completamente esta última frase, que tal usar o ChatGPT para pedir uma explicação mais clara sobre por que a não diferenciabilidade impede o uso da retropropagação?*
- **Sigmoid:** transforma a saída em um valor entre 0 e 1, interpretado como probabilidade. É diferenciável, mas sofre com **saturação**: quando os valores de entrada são muito altos ou muito baixos, a derivada tende a zero, dificultando o aprendizado.
- **Tanh:** semelhante à sigmoid, mas com saída entre  $-1$  e  $+1$ . Também sofre saturação, mas é simétrica em torno da origem, o que a torna mais adequada para camadas intermediárias do que a sigmoid.
- **ReLU (Rectified Linear Unit):** atualmente a mais utilizada em redes profundas. Define saída 0 para valores negativos e saída igual à entrada para valores positivos. É simples, eficiente e evita parte dos problemas de saturação, mas pode gerar “neurônios mortos” (que ficam sempre em 0).
- **Leaky ReLU:** variação da ReLU que permite um pequeno valor negativo para entradas menores que zero, reduzindo o problema dos neurônios mortos.
- **Softmax:** muito usada na camada de saída em problemas de classificação multiclasse, pois transforma o vetor de saídas em uma distribuição de probabilidades (a soma é sempre 1).

### 3. Como interpretar no contexto aplicado

No **supermercado Buonopreço**, podemos pensar nas funções de ativação como diferentes formas de **decidir ou medir o interesse do cliente**:

- Degrau: decisão rígida — envia ou não envia um cupom.
- Sigmoid: grau de probabilidade de que o cliente responda a uma promoção.
- ReLU: intensidade proporcional ao gasto médio ou frequência de compras.

- Softmax: classificação do cliente em um dos segmentos (ex.: “Fiel Econômico”, “Reativo a Promoções”, etc.).

O estudo das funções de ativação permite que você perceba que não existe uma "única função ideal". A escolha depende do **problema**, da **arquitetura da rede** e do **tipo de saída desejada**.

**Agora é sua vez!** Abra o notebook disponível no GitHub [neste link](#) e visualize o gráfico de cada função de ativação. Explore interativamente como cada curva se comporta e reflita por que algumas funções são mais adequadas em certas situações do que em outras.