

## Model

Ótima escolha! A Função Softmax é uma função de ativação muito importante, especialmente na **camada de saída** de redes neurais usadas para **problemas de classificação multi-classe**.

**Propósito da Softmax:**

A Softmax transforma um vetor de valores reais (os "logits" ou "scores" brutos da rede) em uma distribuição de probabilidade. Isso significa que a saída de cada neurônio será um valor entre 0 e 1, e a soma de todas as saídas será igual a 1. Cada valor de saída pode ser interpretado como a probabilidade de a entrada pertencer à classe correspondente.

**A Fórmula da Função Softmax:**

Para um vetor de entrada  $z = [z_1, z_2, \dots, z_K]$ , onde  $K$  é o número de classes, a Softmax calcula a probabilidade para cada classe  $i$  como:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Onde:

- $\sigma(z)_i$  é a probabilidade da  $i$ -ésima classe.
- $e^{z_i}$  é o exponencial da entrada  $z_i$  para a  $i$ -ésima classe.
- $\sum_{j=1}^K e^{z_j}$  é a soma dos exponenciais de todas as entradas (de todas as classes  $j = 1$  até  $K$ ). Este termo no denominador garante que as probabilidades somem 1.

**Características Principais da Softmax:**

- **Saída como Probabilidades:** Garante que as saídas da camada final sejam probabilidades válidas para cada classe.
- **Soma Unitária:** A soma de todas as probabilidades de saída é sempre 1.
- **"Winner Takes All" Implicitamente:** Embora todas as classes recebam uma probabilidade, a classe com o maior "logit" (entrada  $z_i$ ) terá uma probabilidade significativamente maior, tornando-a a previsão mais provável.
- **Sensível a Grandes Diferenças:** O uso da função exponencial amplifica as diferenças entre os valores de entrada, tornando a classe com a entrada mais alta ainda mais dominante na distribuição de probabilidade.

**Exemplo de Cálculo:**

Suponha que temos uma rede neural que está tentando classificar uma imagem em três classes (por exemplo, "gato", "cachorro", "pássaro"). Após a última camada linear, obtemos os seguintes logits (entradas  $z$ ) para essas três classes:

$$z = [z_{\text{gato}}, z_{\text{cachorro}}, z_{\text{pássaro}}] = [2.0, 1.0, 0.5]$$

Vamos aplicar a função Softmax:

## 1. Calcule os exponenciais de cada logit:

- $e^{z_{\text{gato}}} = e^{2.0} \approx 7.389$
- $e^{z_{\text{cachorro}}} = e^{1.0} \approx 2.718$
- $e^{z_{\text{pássaro}}} = e^{0.5} \approx 1.649$

## 2. Calcule a soma dos exponenciais (o denominador):

$$\sum e^{z_j} = 7.389 + 2.718 + 1.649 \approx 11.756$$

## 3. Calcule a probabilidade para cada classe:

- $P(\text{gato}) = \frac{e^{2.0}}{\sum e^{z_j}} = \frac{7.389}{11.756} \approx 0.628$
- $P(\text{cachorro}) = \frac{e^{1.0}}{\sum e^{z_j}} = \frac{2.718}{11.756} \approx 0.231$
- $P(\text{pássaro}) = \frac{e^{0.5}}{\sum e^{z_j}} = \frac{1.649}{11.756} \approx 0.140$

**Resultados:**

As probabilidades de saída seriam aproximadamente:

- Gato: 62.8%
- Cachorro: 23.1%
- Pássaro: 14.0%

Note que a soma das probabilidades é  $0.628 + 0.231 + 0.140 \approx 0.999$ , que é aproximadamente 1 (a pequena diferença é devido ao arredondamento).

Aqui está uma imagem que ilustra o conceito da função Softmax transformando scores brutos em probabilidades:

