

INF721

2023/2



Aprendizado em Redes Neurais Profundas

A3: Regressão Logística

Logística

Avisos

- ▶ Aula A2 - Aprendizado de Máquina publicada no site [slides, vídeo]
- ▶ Aceitar o convite para o Slack!

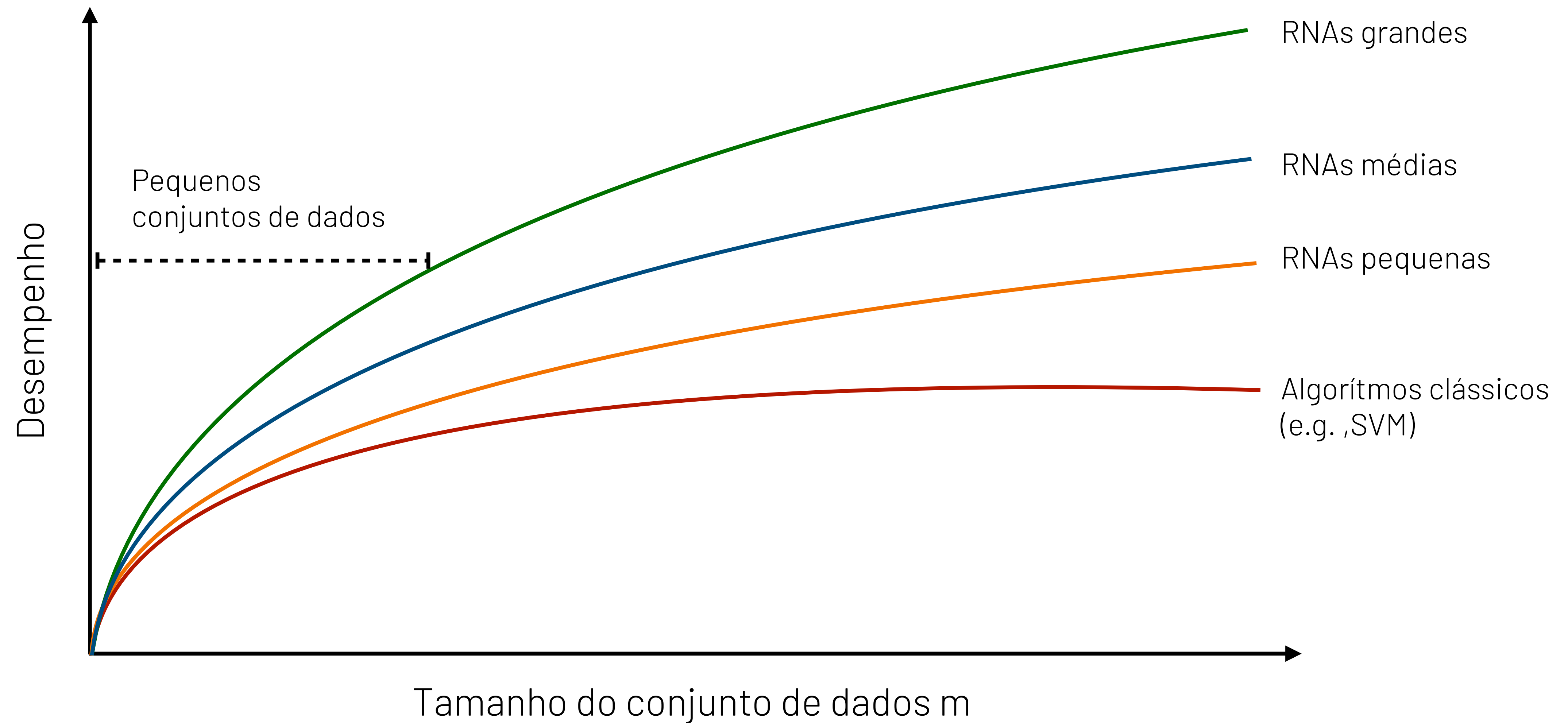
Última aula

- ▶ Tipos de aprendizado e de dados;
- ▶ Espaço de hipóteses e generalização.

Plano de Aula

- ▶ Regressão logística como um neurônio artificial
- ▶ Sucesso das redes neurais artificiais
- ▶ Combinação linear de pesos e entradas
- ▶ Função de ativação
- ▶ Função de perda

O sucesso das Redes Neurais Artificiais (RNAs)

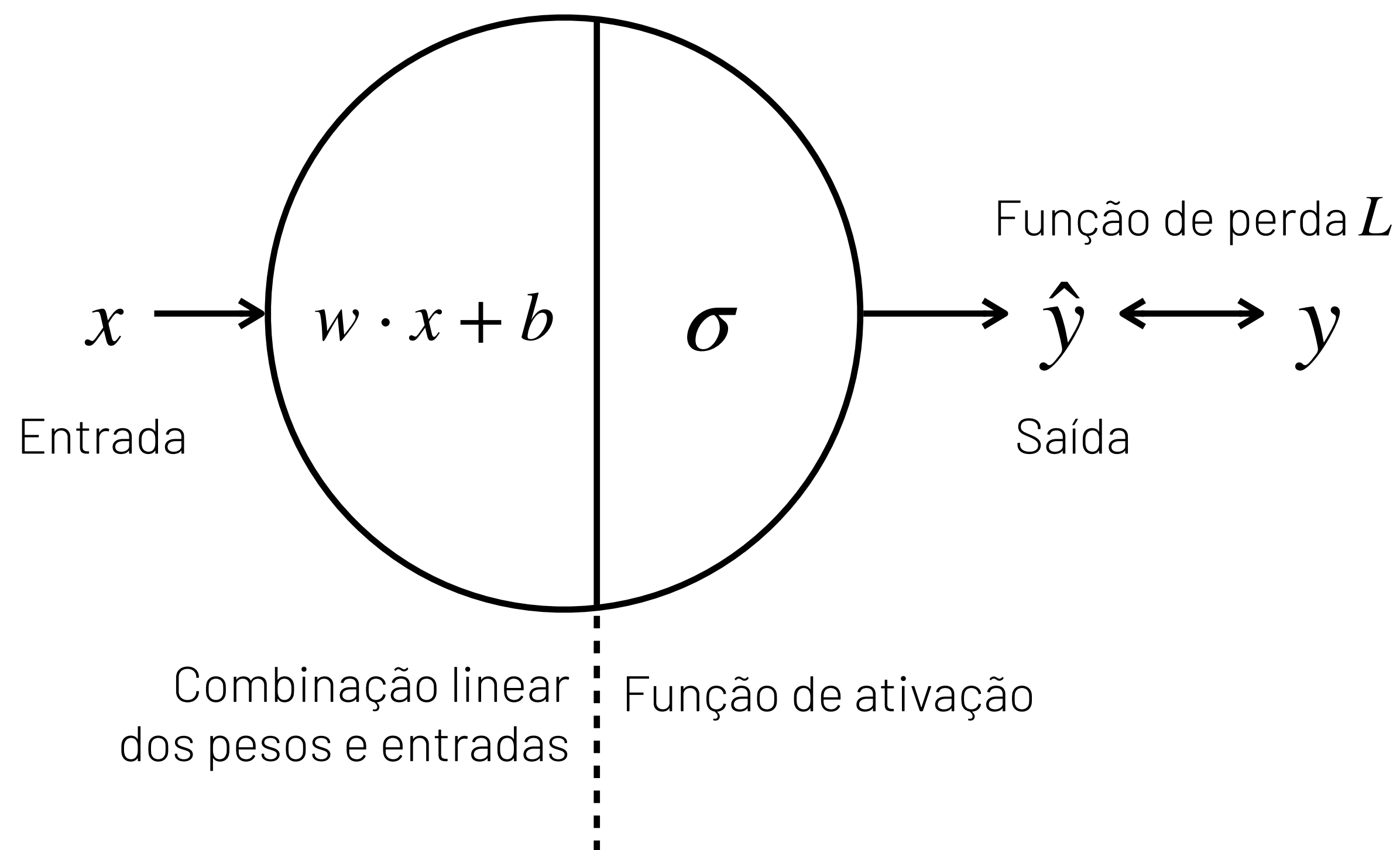


O sucesso das Redes Neurais Artificiais (RNAs)

- ▶ Grandes volumes de dados
- ▶ Computação em GPUs e TPUs
- ▶ Frameworks de derivação automática
- ▶ Novos algoritmos

Regressão Logística

Algoritmo clássico de classificação binária que pode ser visto como um neurônio artificial de uma RNA



Regressão Logística

Entrada

Um exemplo $x \in \mathbb{R}^d$

Saída

A probabilidade de x ser da classe $y = 1$

$$\hat{y} = P(y = 1 | x), 0 \leq \hat{y} \leq 1$$

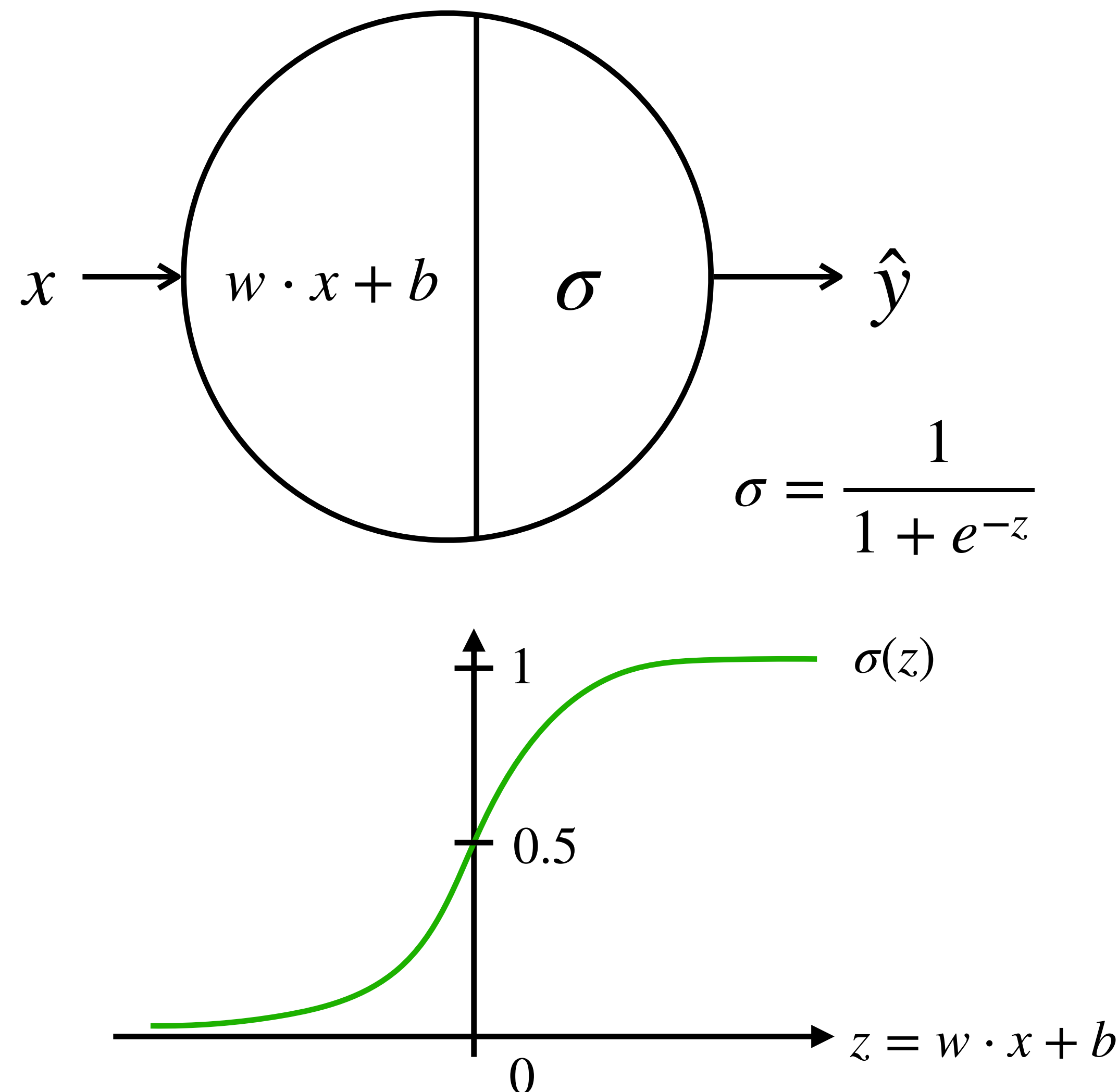
Hipótese

$$h(x) = \sigma(w \cdot x + b)$$

onde:

$w \in \mathbb{R}^d$ e $b \in \mathbb{R}$ são pesos

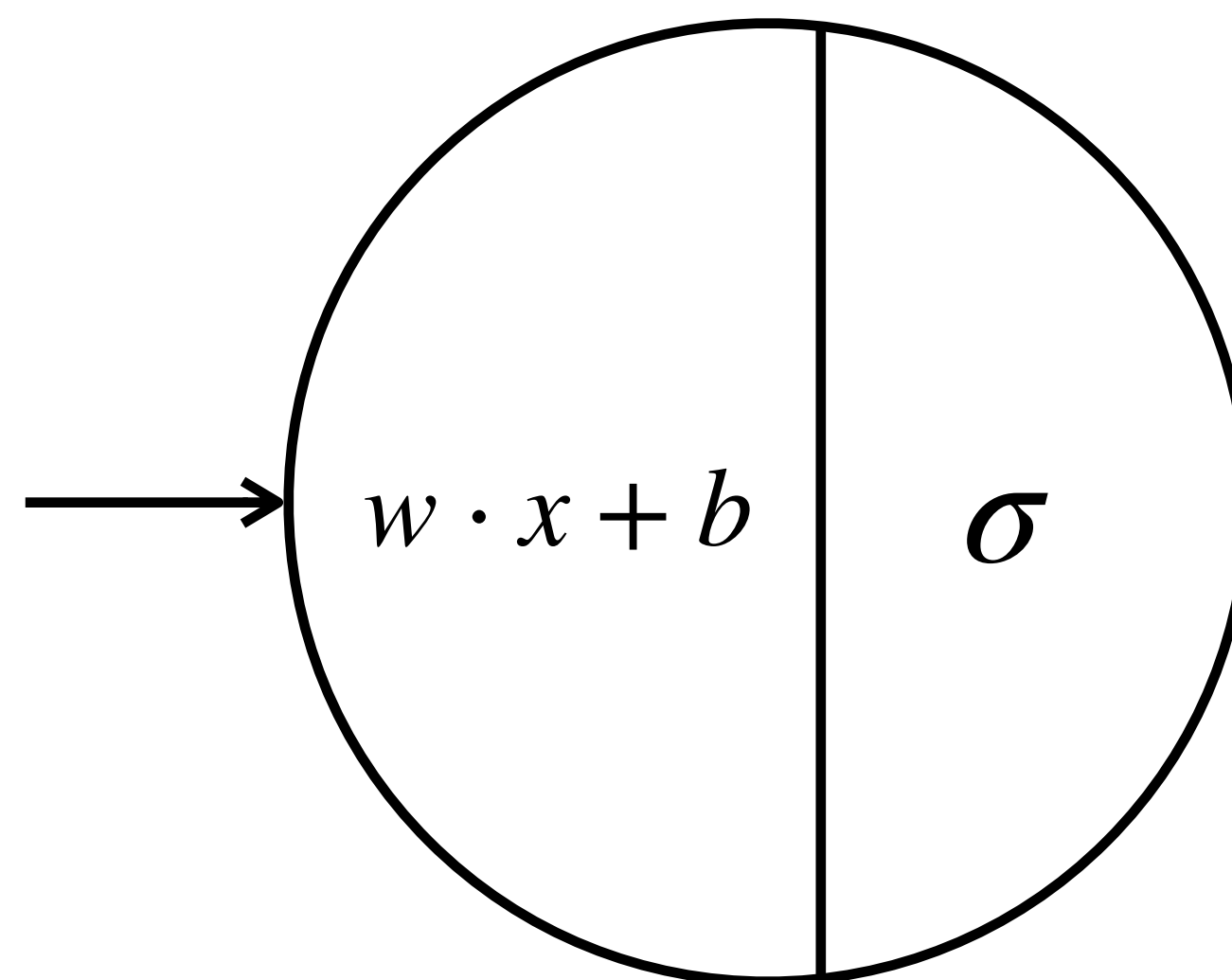
σ é a função logística (sigmoide)



Interpretação de Probabilidade



x



Como os valores da hipótese variam entre 0 e 1, eles podem ser interpretados como valores de probabilidade.

→ 0.8

A imagem x tem 80% de chance de ser da classe Gato (1)

Fronteira de Decisão

Para encontrar o rótulo \hat{y} com a hipótese, usamos um limiar:

$$\hat{y} = \begin{cases} 1, & \text{se } h(x) \geq 0.5 \\ 0, & \text{se } h(x) < 0.5 \end{cases}$$

Repare que o limiar acima é equivalente ao seguinte limiar:

$$\hat{y} = \begin{cases} 1, & \text{se } w \cdot x + b \geq 0 \\ 0, & \text{se } w \cdot x + b < 0 \end{cases}$$

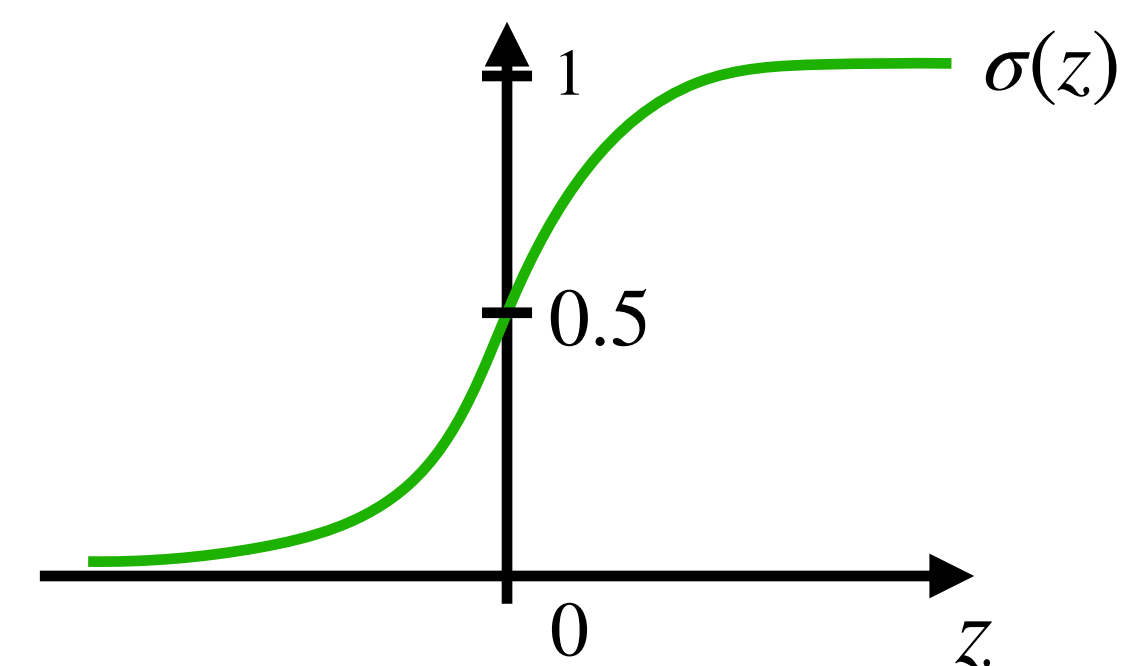
Pois:

$$h(x) \geq 0.5 \text{ quando } z \geq 0$$

$$h(x) < 0.5 \text{ quando } z < 0$$

Regressão Logística

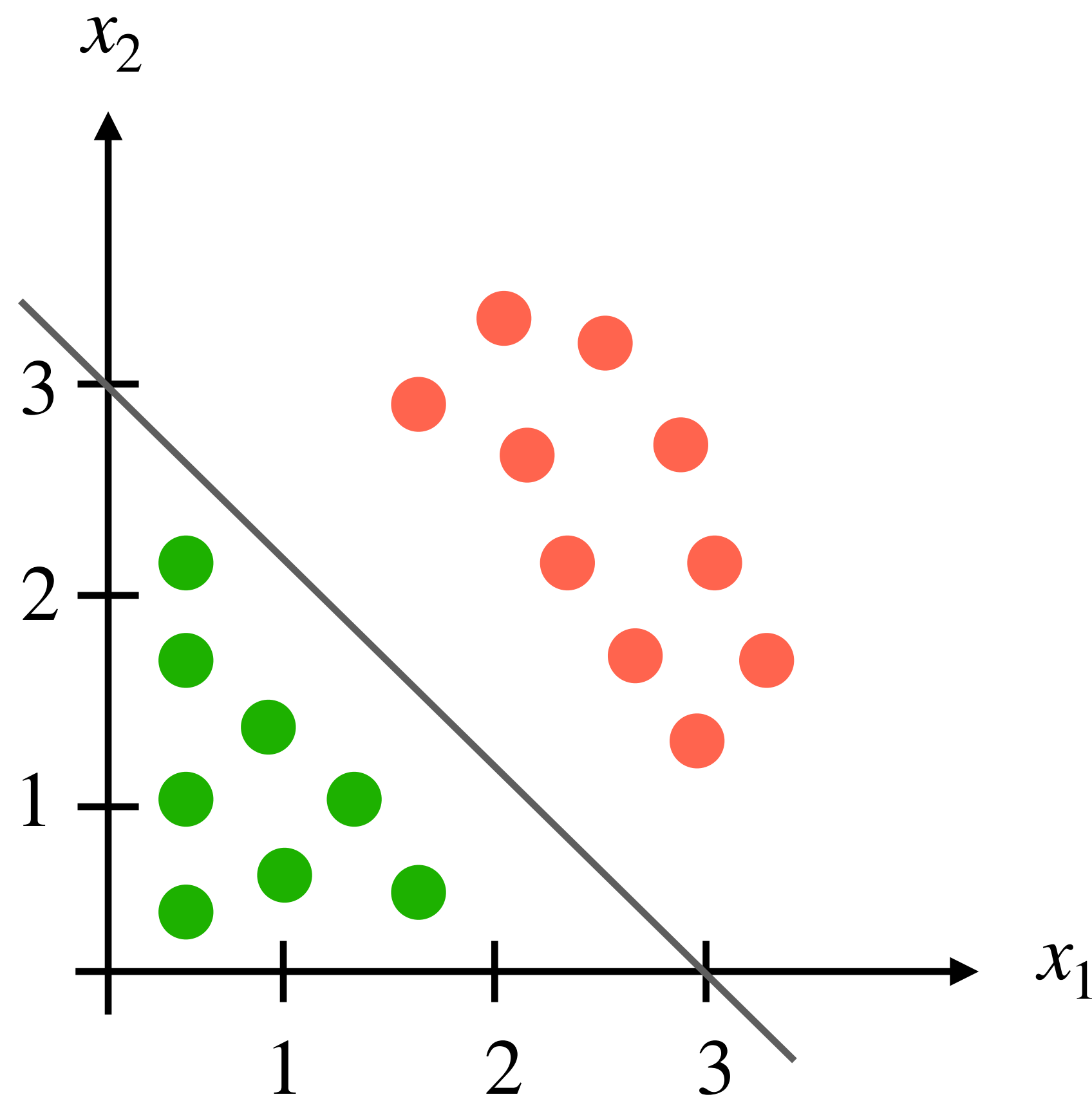
$$h(x) = \frac{1}{1 + e^{-z}}$$



Fronteira de Decisão

Regressão Logística

$$h(x) = \frac{1}{1 + e^{-z}}$$



Considere a seguinte hipótese após o treinamento:

$$w_1 = 1, w_2 = 1, b = -3$$

$$\hat{y} = \begin{cases} 1, & \text{se } x_1 + x_2 - 3 \geq 0 \\ 0, & \text{se } x_1 + x_2 - 3 < 0 \end{cases}$$

A reta $x_1 + x_2 = 3$ é chamada de **fronteira de decisão**.

Função de Perda

Dado um conjunto de dados $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$,
queremos $h(x'_i) \approx y'_i$

Poderíamos tentar usar a função de perda quadrática:

$$L(h) = \frac{1}{n} \sum_{i=1}^n (h(x_i) - y_i)^2$$

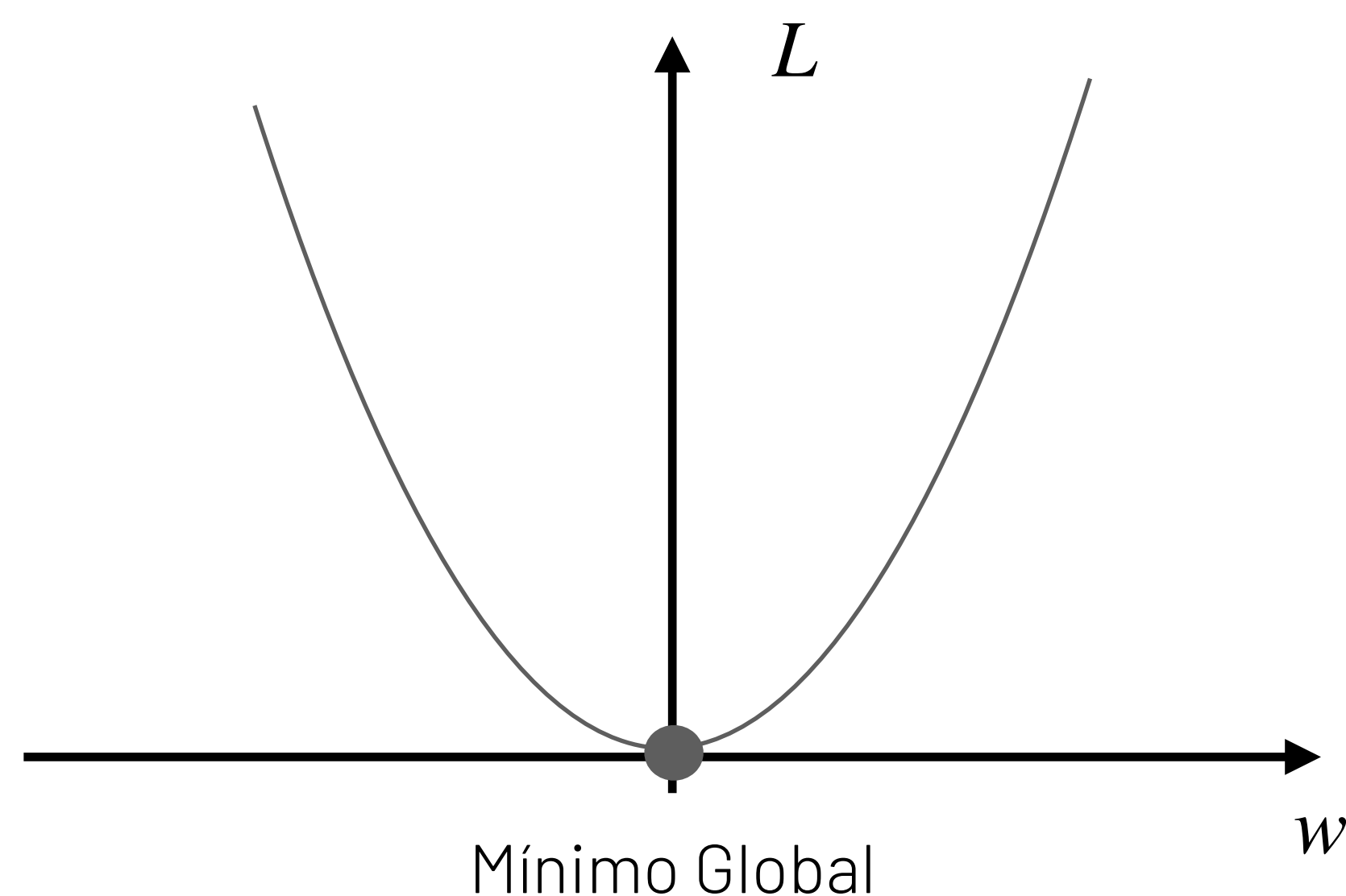
Porém, ela produz um problema de otimização não-convexo!

Regressão Logística

$$h(x) = \frac{1}{1 + e^{-w \cdot x + b}}$$

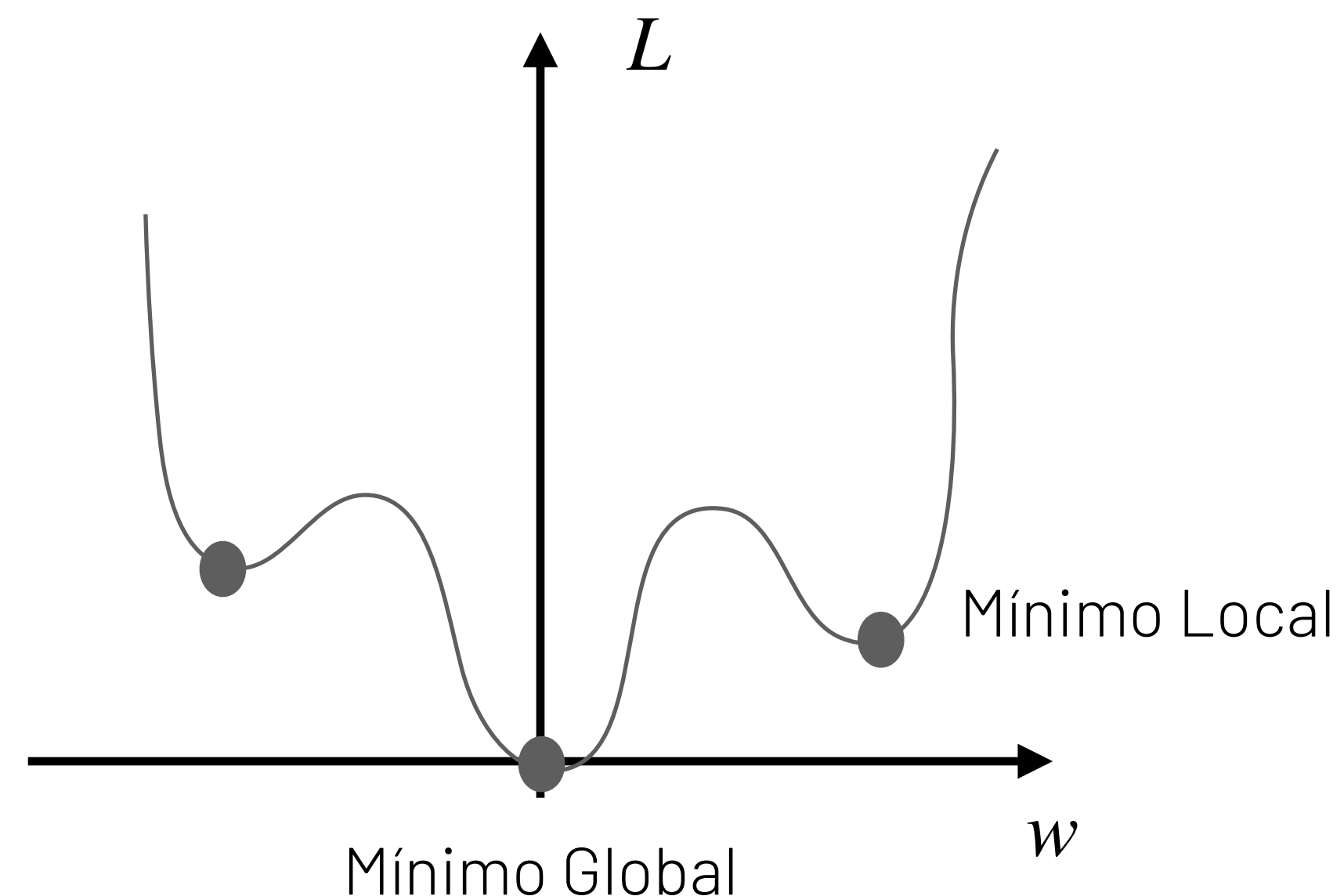
Otimização Convexa

Problema Convexo



Possui apenas um valor mínimo, chamado de mínimo global.

Problema Não-convexo



Além do mínimo global, possui mínimos locais.

Função de Perda

Entropia Cruzada Binária

Dado um conjunto de dados $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$,
queremos $h(x'_i) \approx y'_i$

Em regressão logística, utilizamos a função de entropia cruzada binária como função de perda:

$$L(\hat{y}, y) = - (y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

Que produz um problema de otimização convexo!

Minimar entropia cruzada em problemas de classificação é
equivalente a **estimativa por máxima verossimilhança**
(*maximum likelihood estimation*)!

Regressão Logística

$$h(x) = \frac{1}{1 + e^{-w \cdot x + b}}$$

Função de Perda

Entropia Cruzada Binária

Dado um conjunto de dados $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$,
queremos $h(x'_i) \approx y'_i$

$$P(y = 1 | x) = \hat{y}$$

$$P(y = 0 | x) = 1 - \hat{y}$$

$$P(y | x) = \hat{y}^y (1 - \hat{y})^{(1-y)}$$

$$\log P(y | x) = \log [\hat{y}^y (1 - \hat{y})^{(1-y)}]$$

$$\log P(y | x) = y \log \hat{y} + (1 - y) \log (1 - \hat{y})$$

$$L(\hat{y}, y) = - (y \log \hat{y} + (1 - y) \log (1 - \hat{y}))$$

Regressão Logística

$$h(x) = \frac{1}{1 + e^{-w \cdot x + b}}$$

Próxima aula

A4: Gradiente Descendente

Otimização de pesos da regressão logística com gradiente descendente.