



Regressão logística

Ciência de Dados II

Professor: Gabriel Machado Lunardi
gabriel.lunardi@ufsm.br

Regressão linear...

A regressão linear é uma técnica estatística usada para encontrar **a relação entre variáveis**. Em um contexto de ML, a regressão linear encontra a relação entre **atributos** e **um rótulo**.

Peso do carro – libras em 1.000 (atributo)	Milhas por galão (rótulo)
3.5	18
3,69	15
3,44	18
3,43	16
4,34	15
4,42	14
2,37	24

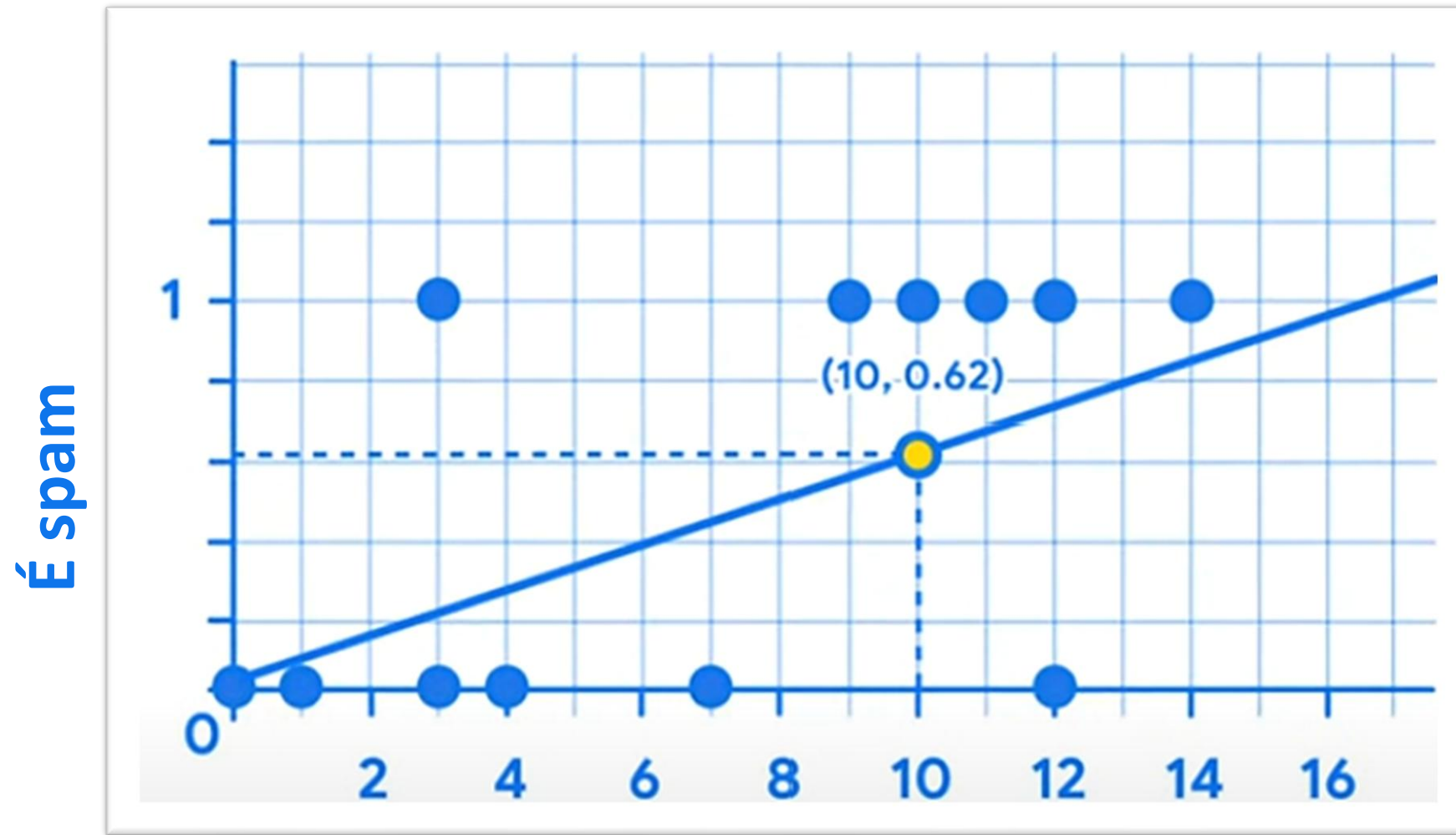
prever a eficiência de combustível de um carro em milhas por galão com base no peso do carro



MAS E SE VOCÊ QUISER CRIAR UM
MODELO PARA RESPONDER A
PERGUNTAS COMO

“VAI CHOVER HOJE?”
“ESSE E-MAIL É SPAM?”

É SPAM / NÃO É SPAM



1 = Spam

0 = Not Spam

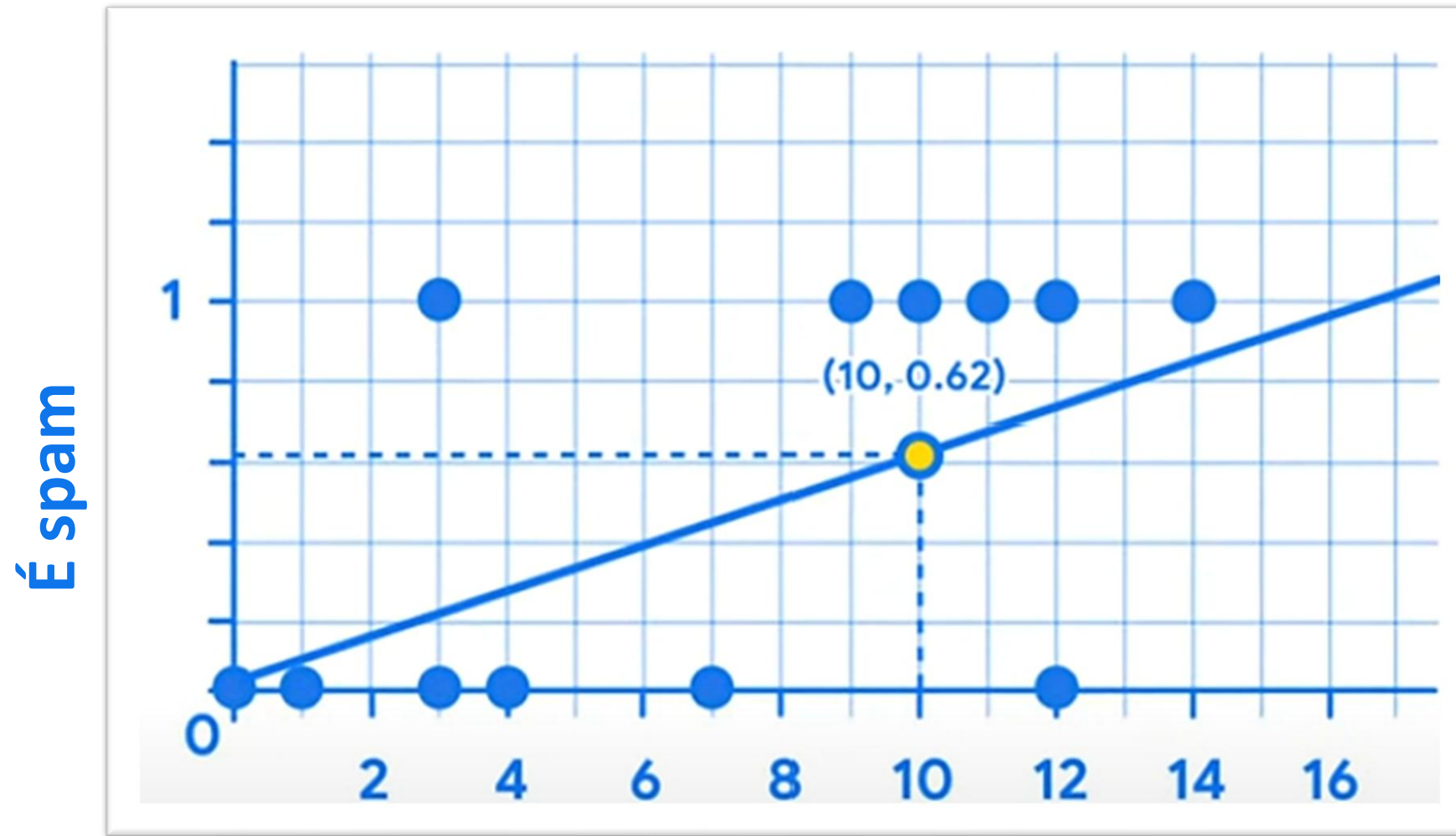
0.62 = ???

Contagem de palavras spam

É SPAM / NÃO É SPAM



Precisamos de um método capaz de gerar essas probabilidades



1 = Spam

0 = Not Spam

0.62 = ???

Contagem de palavras spam

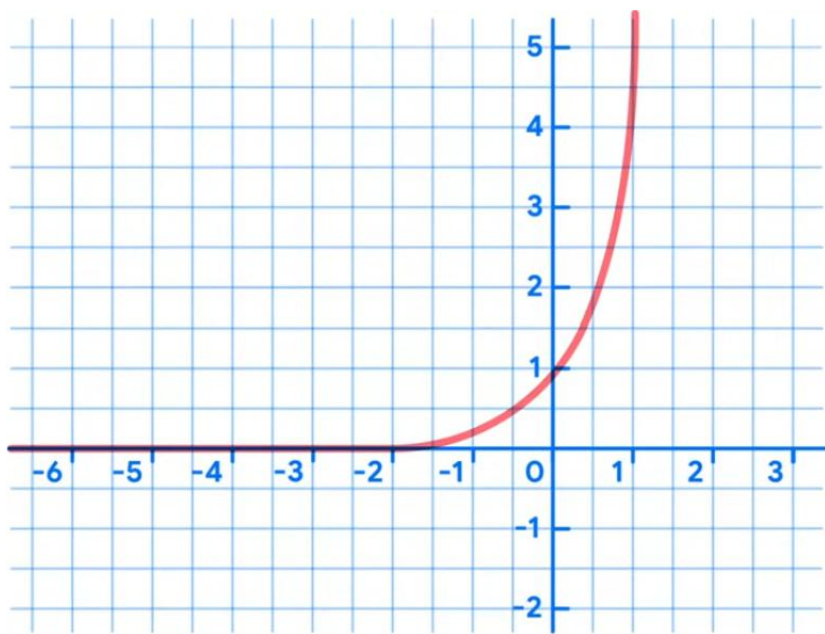
Regressão logística

É um tipo de modelo de regressão que prevê uma probabilidade, isto é, um valor que vai de **0** a **1**, sendo **0%** de chance de algo acontecer e **100%** de chance de algo acontecer.

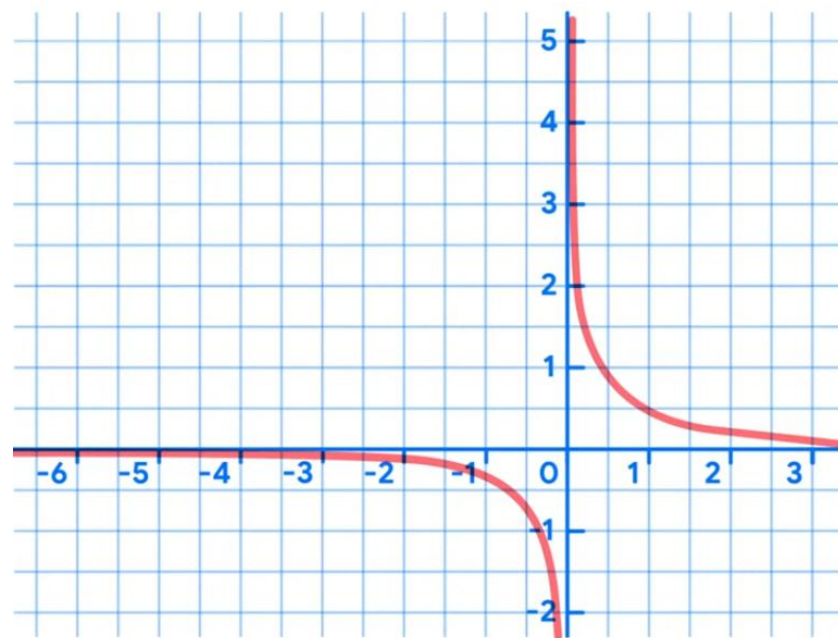
Regressão logística

É um tipo de modelo de regressão que prevê uma probabilidade, isto é, um valor que vai de **0** a **1**, sendo **0%** de chance de algo acontecer e **100%** de chance de algo acontecer.

Duas curvas matemáticas nos permitem estabelecer esses limites!



$$g(x) = e^x$$

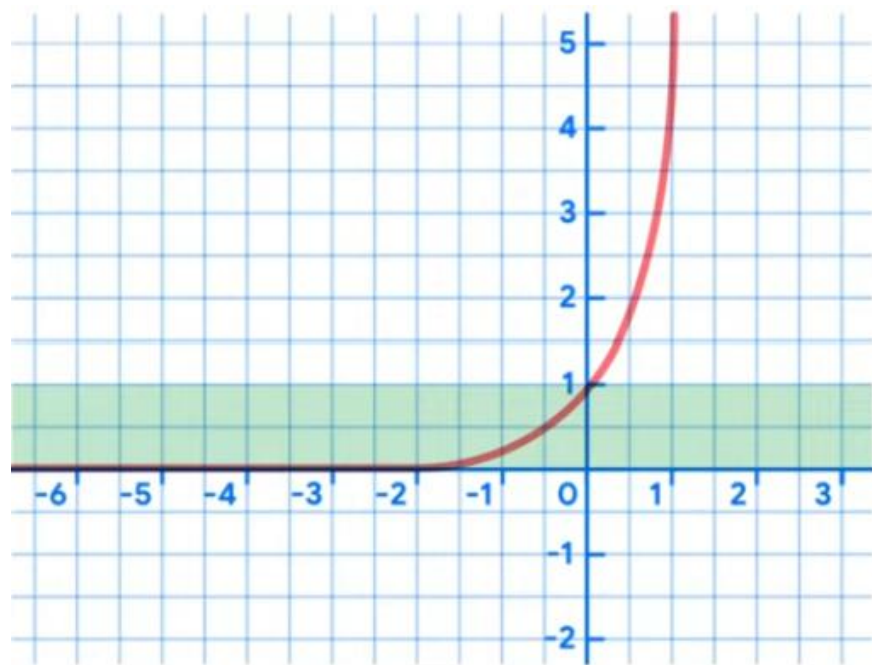


$$g(x) = 1/x$$

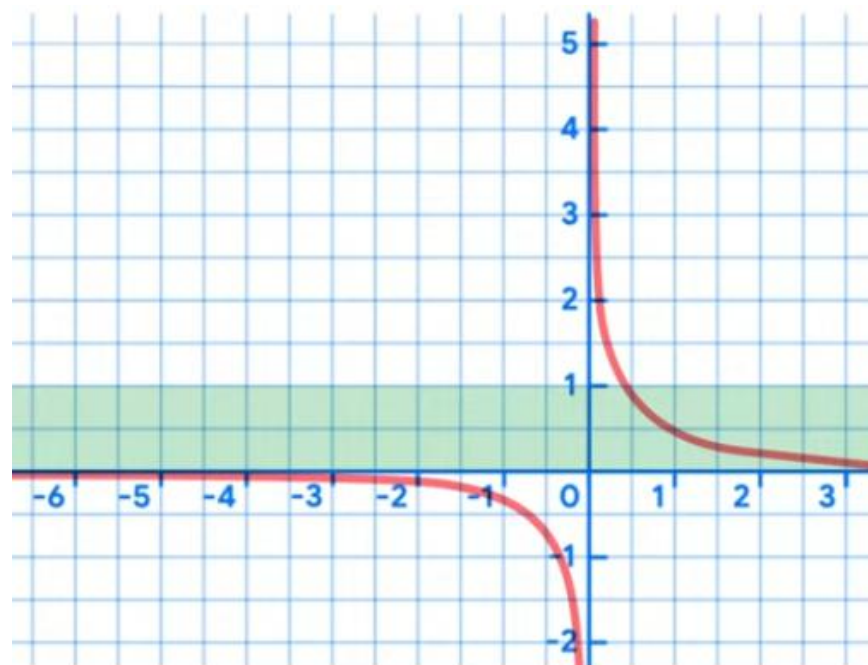
Regressão logística

É um tipo de modelo de regressão que prevê uma probabilidade, isto é, um valor que vai de **0** a **1**, sendo **0%** de chance de algo acontecer e **100%** de chance de algo acontecer.

Duas curvas matemáticas nos permitem estabelecer esses limites!



$$g(x) = e^x$$



$$g(x) = 1/x$$

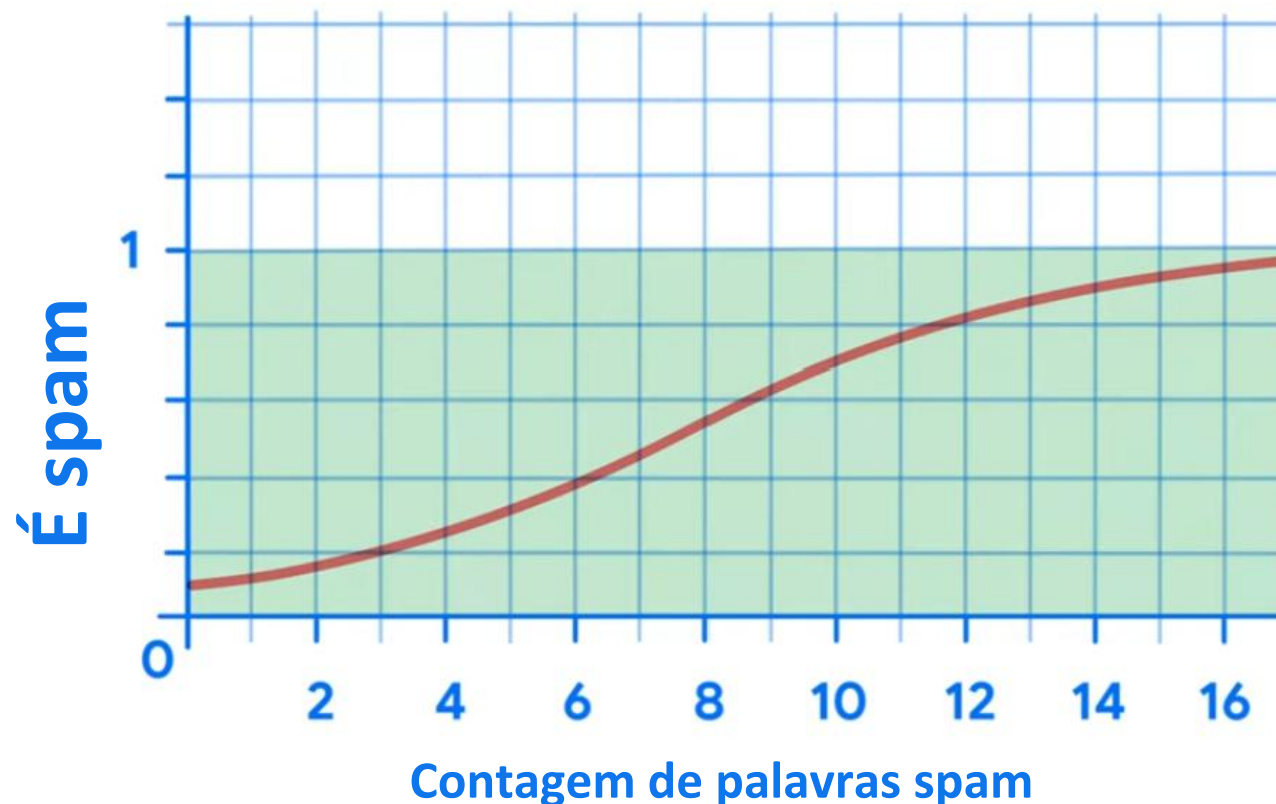


Mas não sozinhas!

Regressão logística

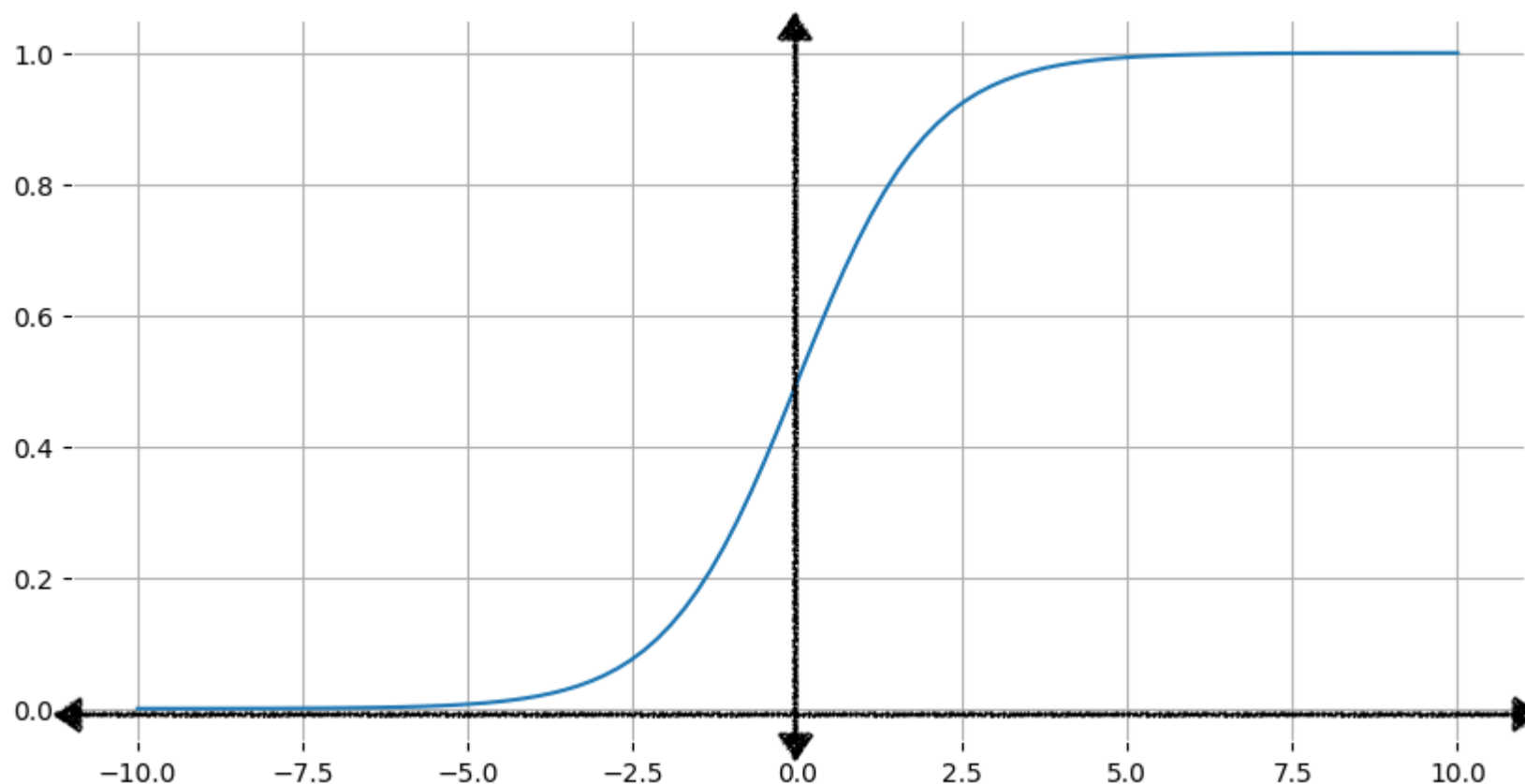
É um tipo de modelo de regressão que prevê uma probabilidade, isto é, um valor que vai de **0** a **1**, sendo **0%** de chance de algo acontecer e **100%** de chance de algo acontecer.

A combinação delas resulta na curva “sigmóide” S que caracteriza a reg. log.



Regressão logística

Um exemplo mais matemático: quanto mais a função tende ao infinito positivo, mais próximo de 100%. Quanto mais a função tende ao infinito negativo, mais próximo de 0%.



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

O componente linear da regressão logística

$$z = b + w_1 x_1 + w_2 x_2 + \dots + w_N x_N$$

Z é a saída da equação linear, também chamada de log probabilidades

b é a polaridade (viés)

✓ É um parâmetro do modelo que é calculado durante o treinamento.

w1, w2 é o **peso** do elemento.

✓ É um parâmetro do modelo que também é calculado durante o treinamento.

x1, x2, ... são **atributos**, ou seja, **entradas**.

A previsão da regressão logística

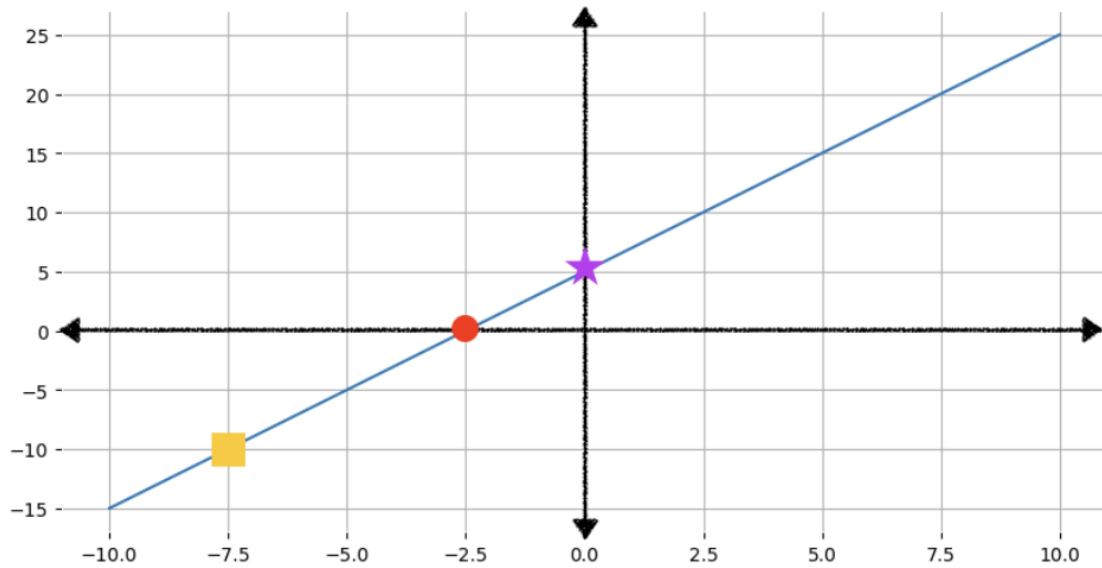
$$y' = \frac{1}{1 + e^{-z}}$$

y' é a saída do modelo (valor de probab. previsto).

z é a saída linear (dada pelo cálculo do slide anterior).

Exemplo

$$z = 2x + 5$$



$$y' = 1 / (1 + e^{-z})$$

