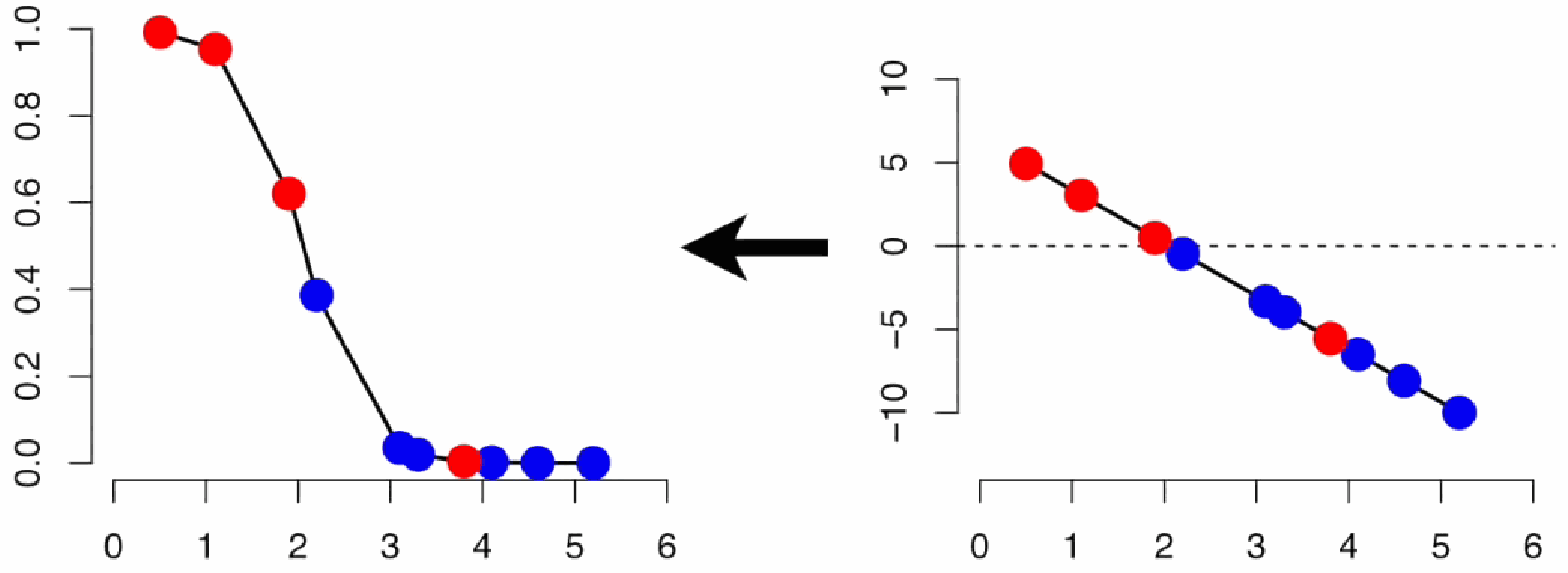
# Regressão Logística

Prof Dr. Vladimir C. Alencar
LANA/UEPB
www.valencar.com

LANA
DATA ANALYSIS LABORATORY

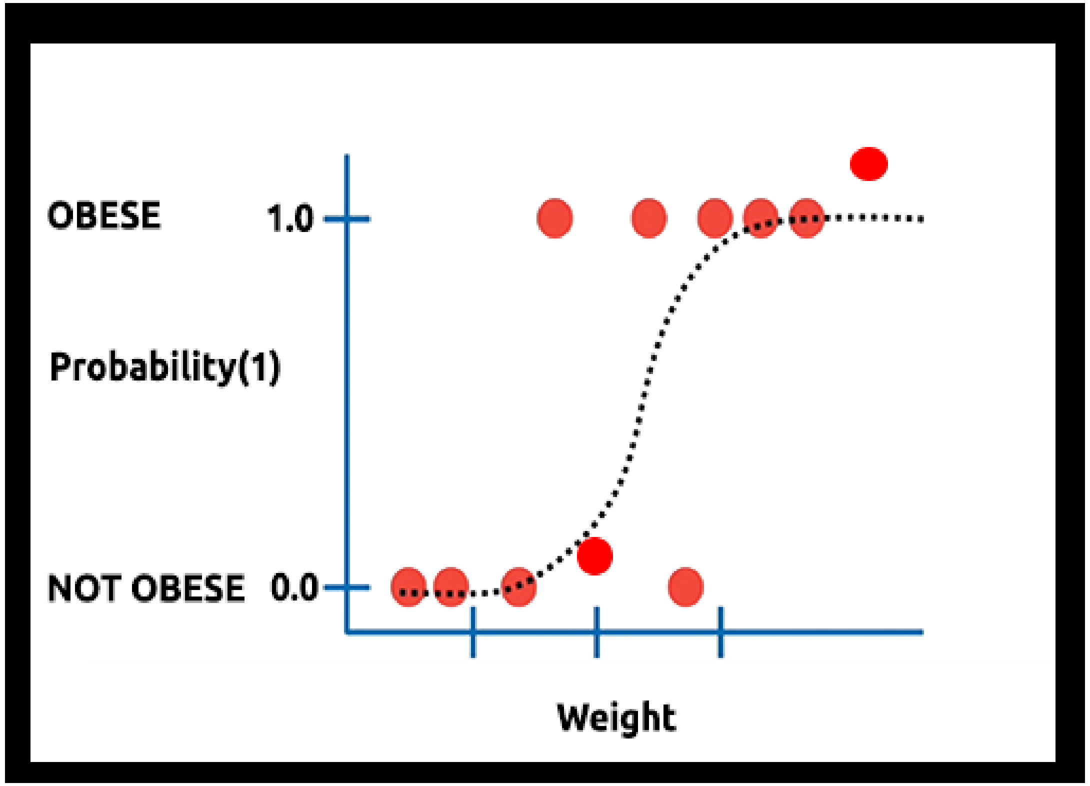
# Regressão Logística



# Regressão Logística

- É um método estatístico usado para modelar a relação entre uma variável dependente binária (ou categórica) e uma ou mais variáveis independentes (preditoras).

- Embora o nome sugira, a regressão logística não é usada para previsões contínuas, mas sim para estimar a probabilidade de um determinado evento ocorrer, ou seja, é especialmente útil para classificação.

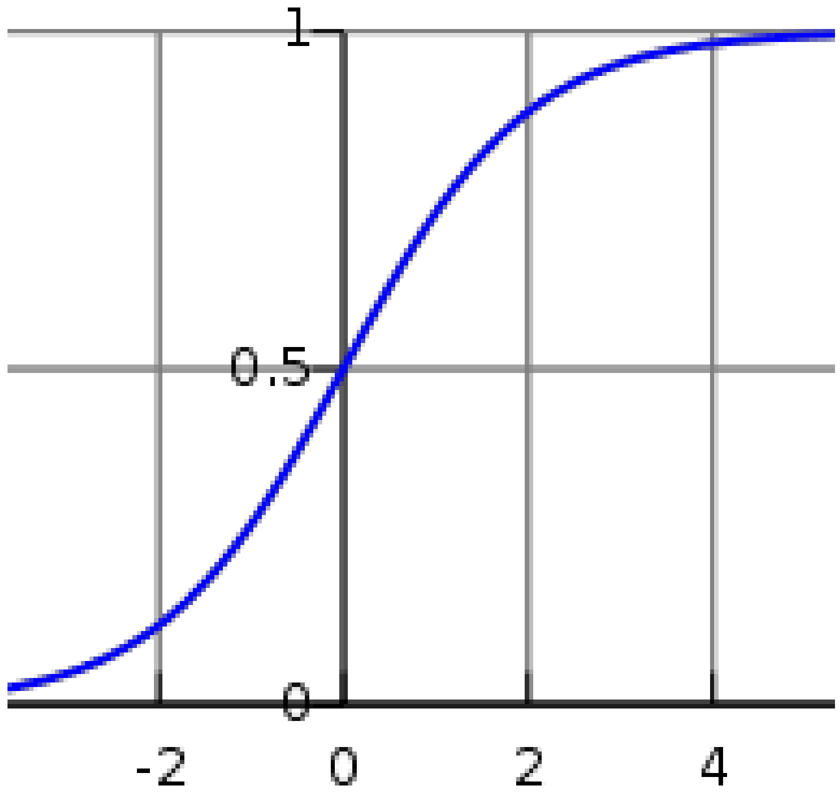


# Regressão Logística

- A regressão logística é um modelo de classificação que trabalha muito bem com classes linearmente separáveis.
- É um dos algoritmos de classificação mais amplarmente utilizados na indústria.
- É um modelo de classificação binária que pode ser estendido para mutliclasse.

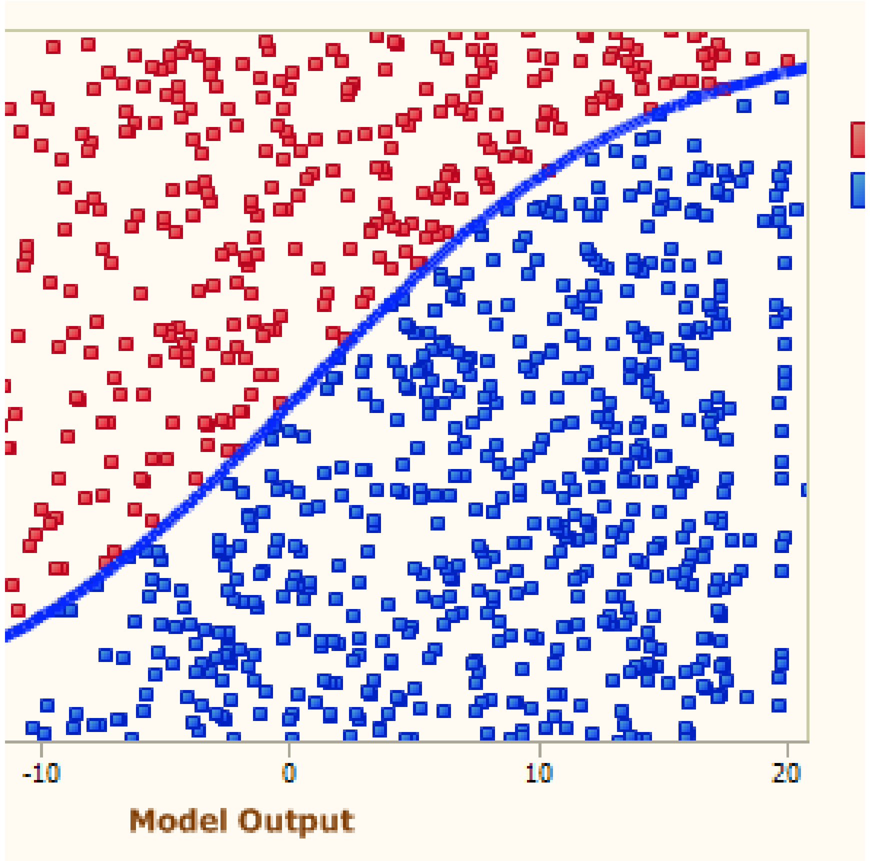
# Regressão Logística

- A regressão logística é uma técnica estatística que tem como objetivo modelar, a partir de um conjunto de observações, a relação "logística" entre uma variável resposta e uma série de variáveis explicativas numéricas (contínuas, discretas) e/ou categóricas.



# Regressão Logística

- A regressão logística é amplamente usada em ciências médicas e sociais, e tem outras denominações,
  como **modelo logístico**, **modelo logit**, e **classificador de máxima entropia.**



# Regressão Logística

- O modelo tem um comportamento probalístico.
- Trabalha com a probabilidade relacionada ao sucesso.
- Um evento de sucesso não necessariamente significa um evento bom, mas um evento que desejamos prever.
- Ex. A probabilidade que um paciente tenha certa doença

O evento positivo é relacionado com um Rótulo ou Classe

(1 – Tem doença, 0 – Não tem)

- Para mapearmos os eventos, usamos uma função logística (usa o logaritmo da probabilidade do evento).

# Regressão Logística

- **Na Regressão Logística, a variável resposta é binária**

1 - acontecimento de interesse (sucesso)

0 - acontecimento complementar (insucesso)

# Regressão Logística - Formulação

Tranformação Logit

Considere o seguinte exemplo: uma organização quer determinar o aumento salarial de um funcionário com base em seu desempenho

# Regressão Logística

Tranformação Logit

Agora, e se a organização quiser saber se um funcionário obteria uma promoção ou não com base em seu desempenho?

O gráfico linear acima não será adequado neste caso.
Como tal, cortamos a linha em zero e um, e a convertemos em uma curva sigmoide (curva S).


# Regressão Logística

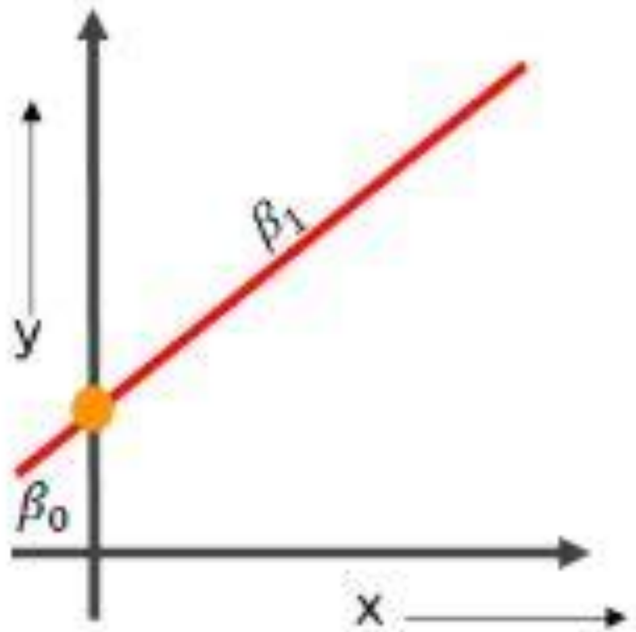


Para entender a regressão logística, vamos analisar as probabilidades de sucesso.

Probabilidades ($\theta$) = Probabilidade de um evento acontecer / Probabilidade de um evento não acontecer
$\theta$ = p / 1 – p

Os valores das probabilidades variam de zero a ∞ e os valores da probabilidade estão entre zero e um.

Considere a equação de uma linha reta:
$y = \beta 0 + \beta 1 * x$



# Regressão Logística

Tranformação Logit

Agora, para prever as probabilidades de sucesso, usamos a seguinte fórmula:

$$\log\left(\frac{p(x)}{1-P(x)}\right) = \beta_0 + \beta_1 x$$

log (Probabiliade) = Valores entre 0 e 1

Exponenciando ambos os lados, temos:

$$e^{ln\left(\frac{p(x)}{1-p(x)}\right)} = e^{\beta_0+\beta_1 x}$$

$$\left(\frac{p(x)}{1-p(x)}\right) = e^{\beta_0+\beta_1 x}$$

# Regressão Logística



Seja Y = e $^{\beta 0+\beta 1 * x}$

Então p(x) / 1 - p(x) = Y

p(x) = Y(1 - p(x))

p(x) = Y - Y(p(x))

p(x) + Y(p(x)) = Y

p(x)(1+Y) = Y

p(x) = Y / 1+Y

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$

A equação da função sigmóide é:

$$p(x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}}$$

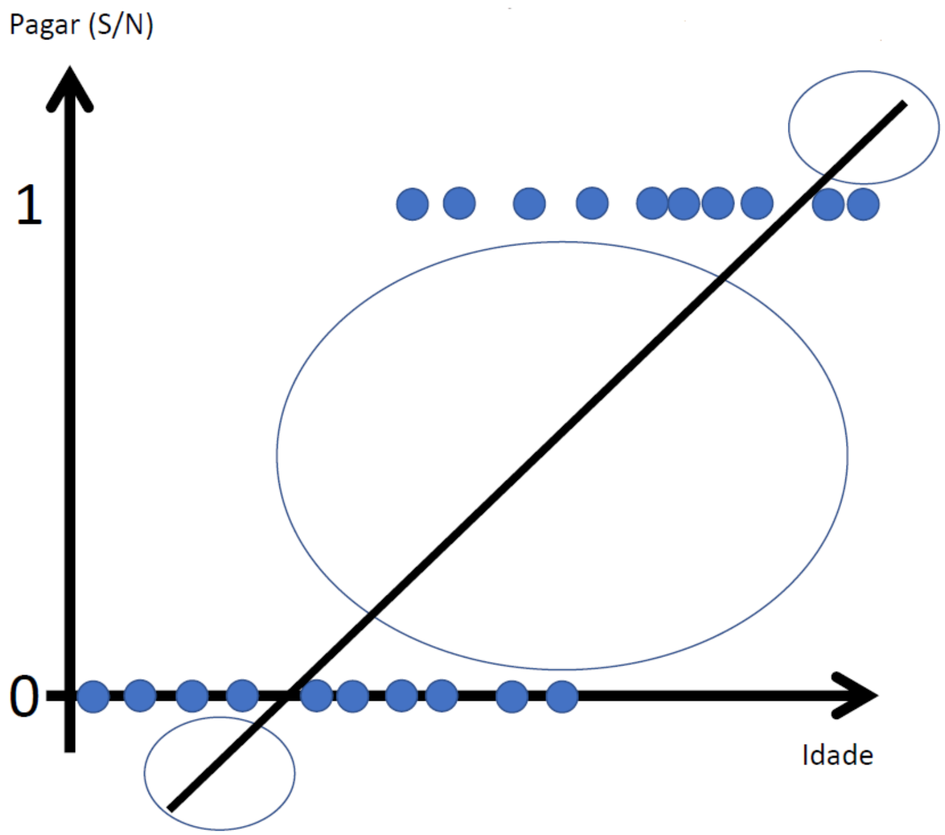
# Regressão Logística



A curva sigmóide obtida a partir da equação acima é a seguinte:

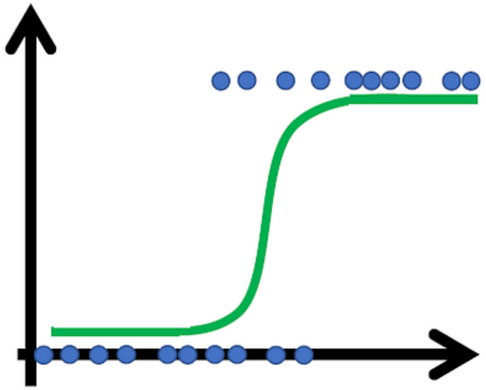

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

# Regressão Logística



# Regressão Logística

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



- Função Sigmóide

$$z = \boldsymbol{w}^T \boldsymbol{x} = w_0 x_0 + w_1 x_1 + \cdots + w_m x_m$$

# Regressão Logística

- O log se refere ao logaritmo natural.
- A função logit trabalha com valores na faixa de 0 a 1 e os transforma em números reais, nos quais expressam o relacionamento linear entre valores e suas probabilidades logarítmicas

$$logit\big(p(y=1|\boldsymbol{x})\big) = w_0x_0 + w_1x_1 + \cdots + w_mx_m = \sum_{i=0}^{m} w_ix_i = \boldsymbol{w}^T\boldsymbol{x}$$

Onde:
**p(y=1|x)** é a probabilidade condicional de um particular evento onde a classe pertence a 1 (sucesso) dado os atributos de x

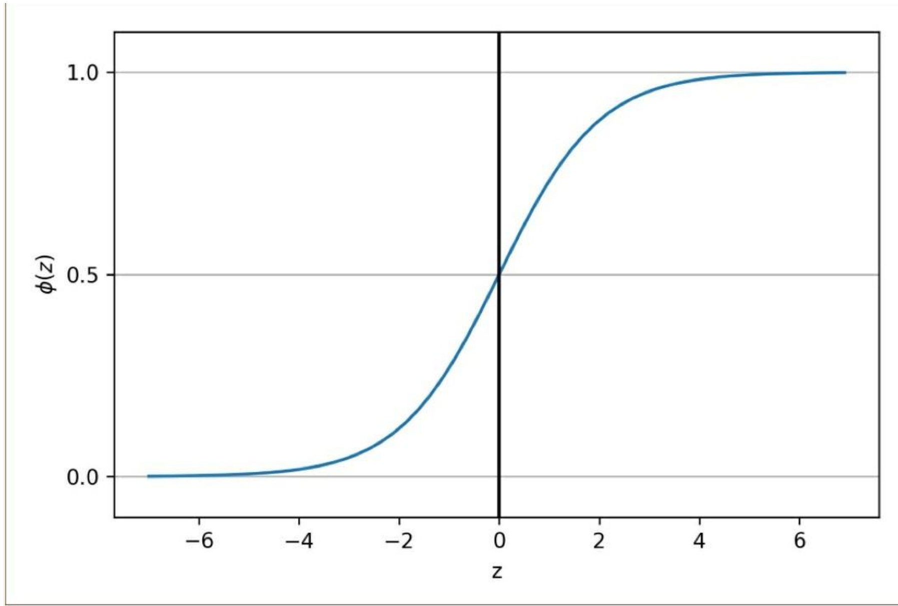
# Regressão Logística

- Nós estamos interessados em prever a probabilidade de certos dados pertencerem a uma classe em particular, então utilizaremos a forma inversa da função logit.
- Nós também chamamos de função logística sigmóide ou função sigmóide, e tem um comportamento em forma de S.

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

$$z = \boldsymbol{w}^T \boldsymbol{x} = w_0 x_0 + w_1 x_1 + \cdots + w_m x_m$$

Onde Z são os dados de entrada, uma combinação linear dos pesos e dos valores dos atributos



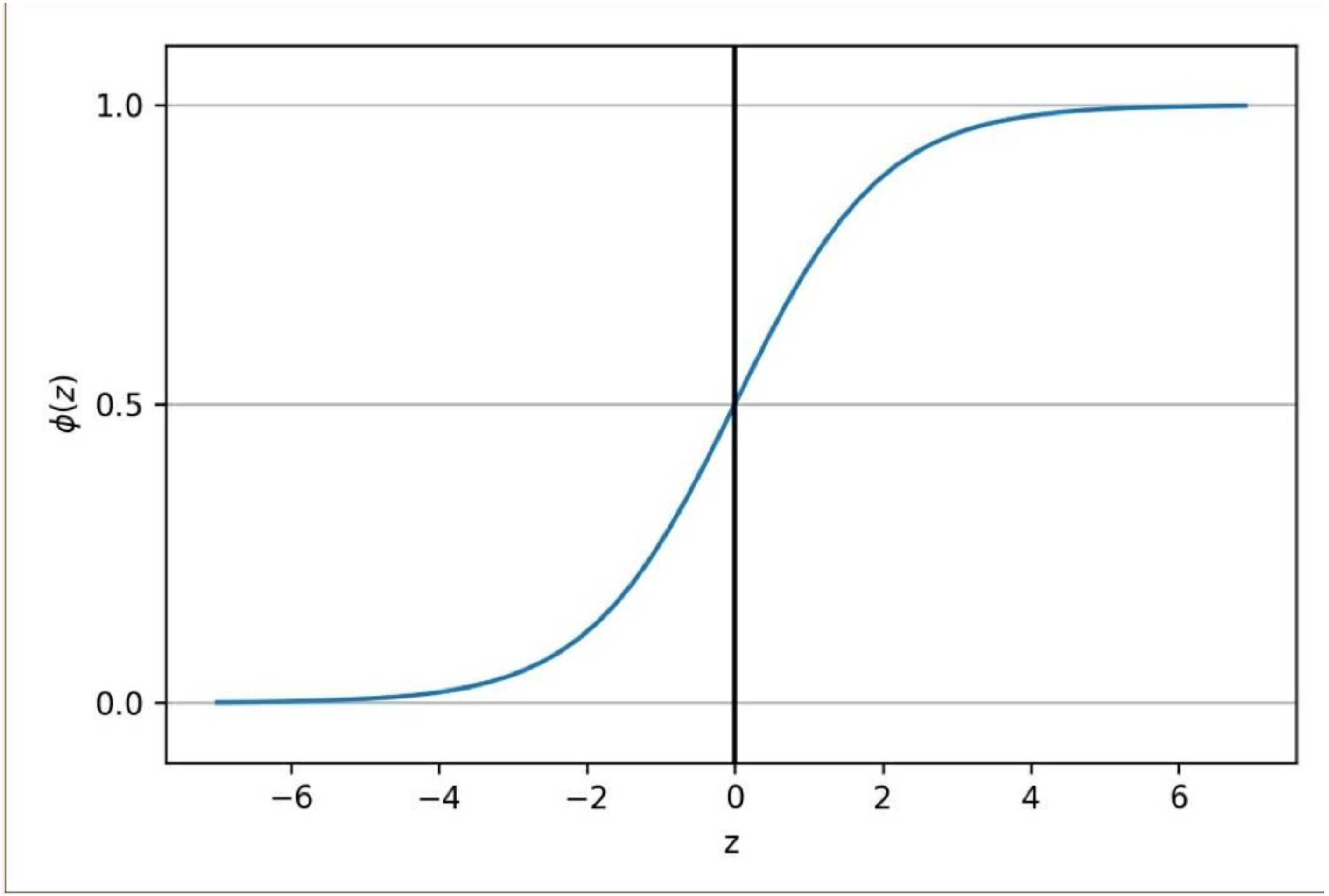
# Regressão Logística

- Função Sigmóide

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

$$z = \boldsymbol{w}^T \boldsymbol{x} = w_0 x_0 + w_1 x_1 + \cdots + w_m x_m$$

e= 2,718281



# Regressão Logística

- A saída da função sigmoid pode ser interpretada com a probabilidade de uma amostra particular pertencer a classe 1, dado seus atributos xi, parametrizados pelos pesos w.

$$\phi(z) = P(y = 1 \mid \boldsymbol{x}; \boldsymbol{w})$$

- Ex. $Se\ \emptyset(z) = 0.8$, para uma detemindada doença, Isto significa 80% de probabilidade de ter a doença (y=1).

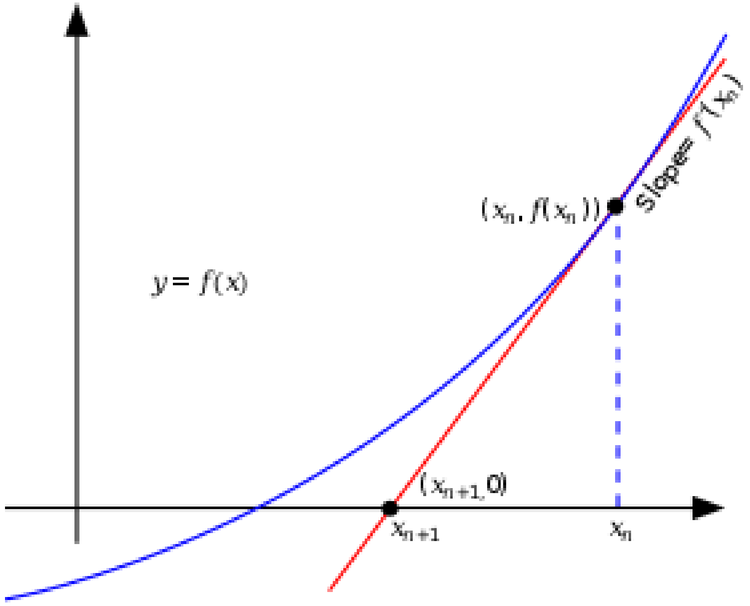
$$P(y = 0 \mid \boldsymbol{x}; \boldsymbol{w}) = 1 - P(y = 1 \mid \boldsymbol{x}; \boldsymbol{w}) = 0.2$$

Significa a probabilidade de que 20% não tem a doença (y=0)

# Otimizador (achar os pesos w)

O método de Newton, é um algoritmo de localização de raízes que produz aproximações sucessivamente melhores para as raízes (ou zeros) de uma função de valor real

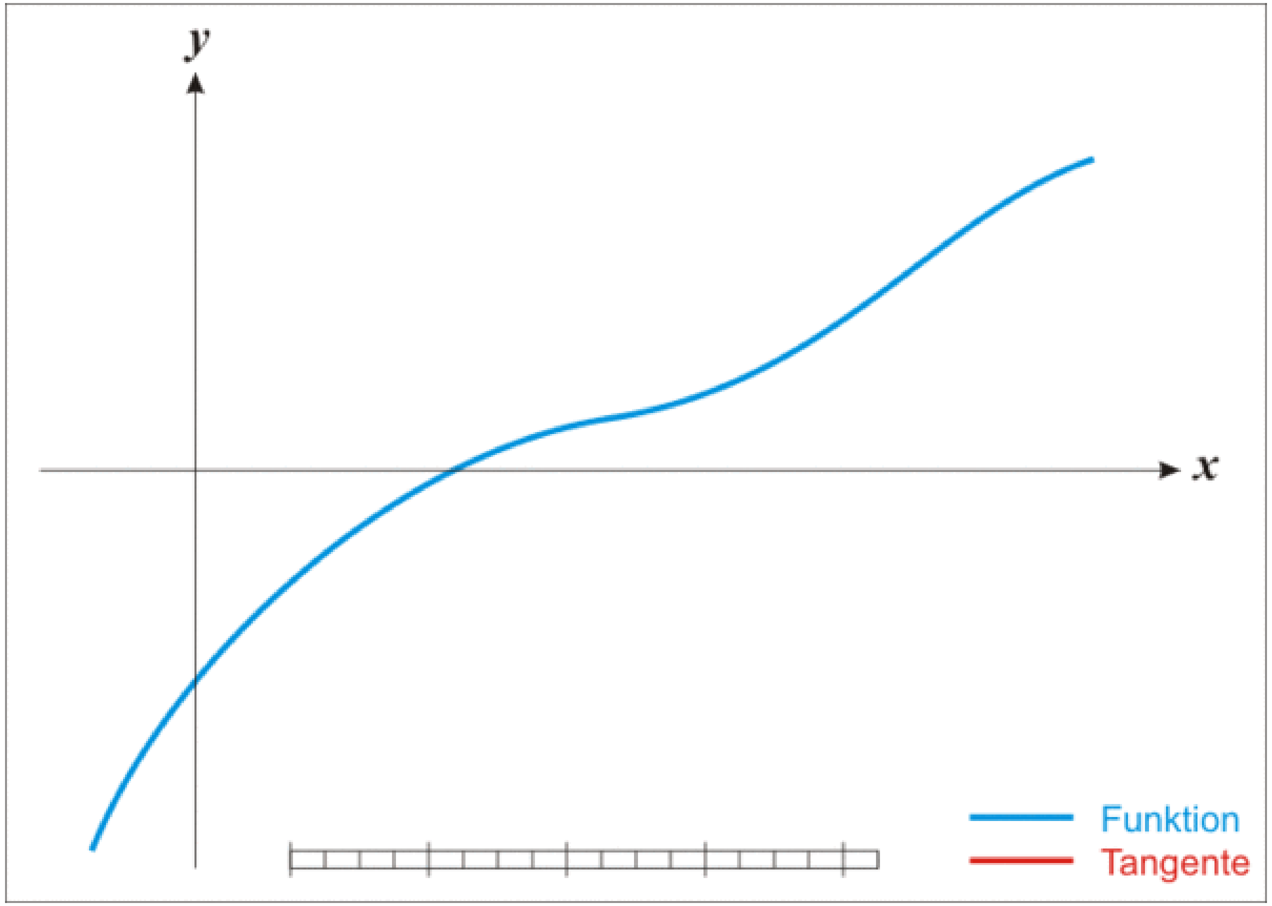
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



# Otimizador (achar os pesos w)

O método de Newton

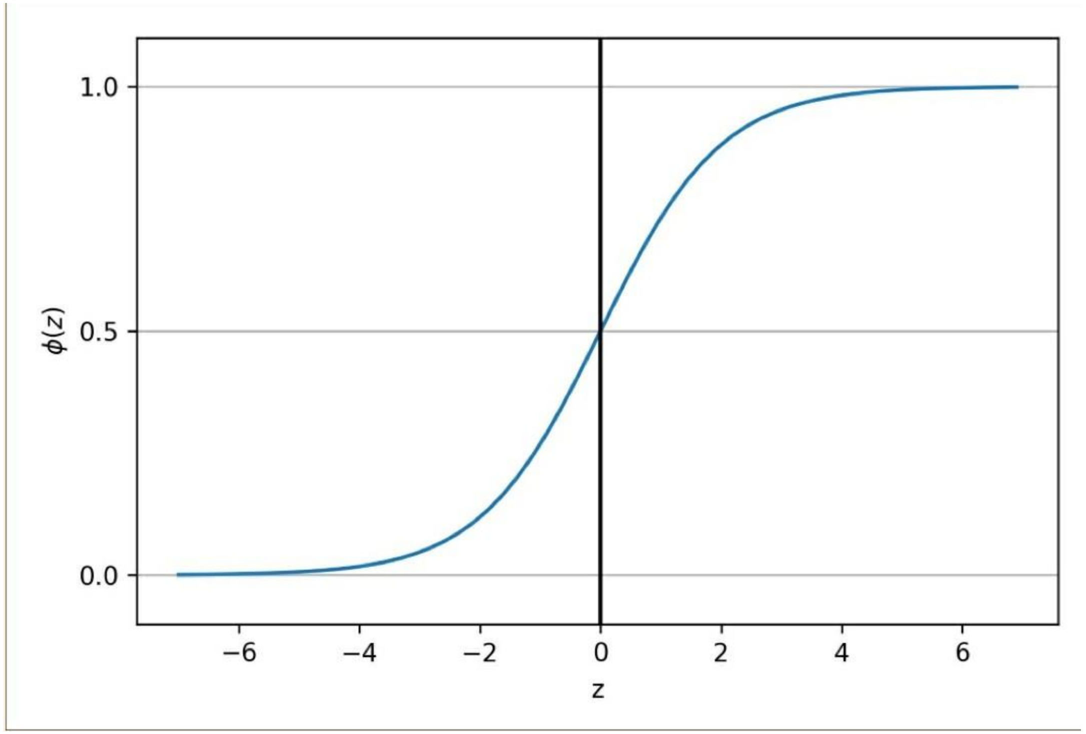
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



# Regressão Logística

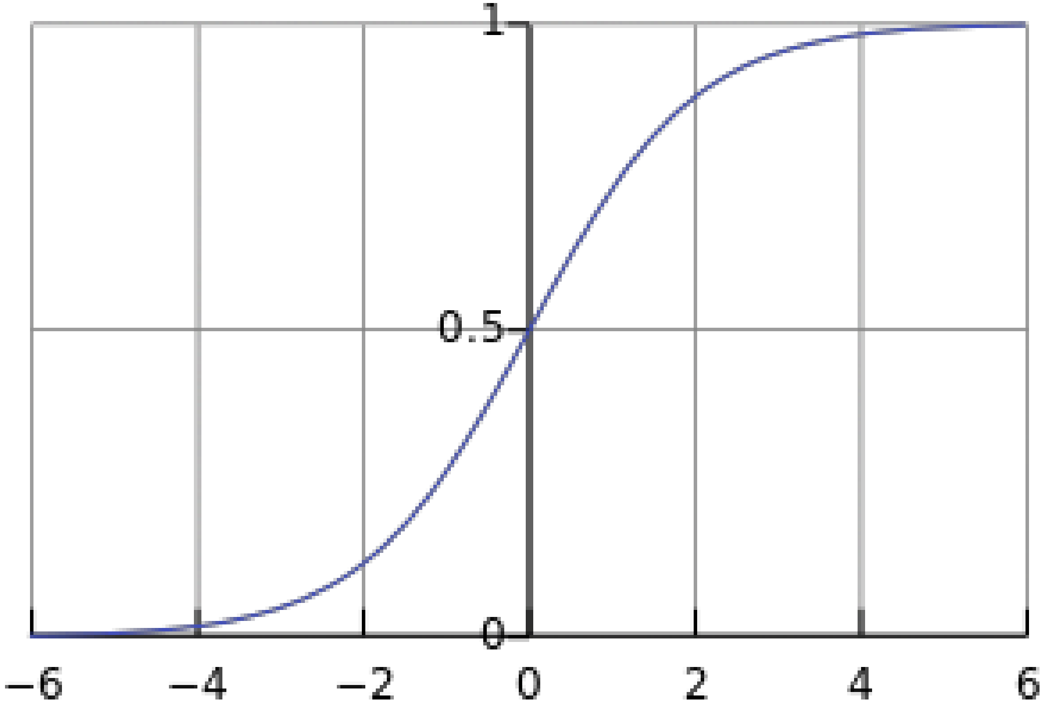
A probilidade prevista será :

$$\hat{y} = \begin{cases} 1 & if\, \phi(z) \geq 0.5 \\ 0 & otherwise \end{cases}$$



# Regressão Logística

- Regressão Logística é útil para modelar
  a probabilidade de um evento ocorrer como função de outros fatores.
- É um modelo linear generalizado que usa como função de ligação a função logit.



# Regressão Logística - utilização

- Diagnóstico médico
- Previsões climáticas
- Segmentação e categorização de imagens
- Reconhecimento de escrita/caligrafia
- Previsão de tempo, não só para prever se vai chover em determindado dia, mas também a chance de chover
- Prever a chance que um paciente ter uma certa doença, dados os sintomas. Um campo onde  a regressão logística é muito popular, que é a medicina.
- Também, em medicina, permite determinar os fatores que caracterizam um grupo de indivíduos doentes em relação a indivíduos saudáveis.
- Na área de seguros, permite encontrar frações de clientes que sejam sensíveis a determinada política securitária em relação a um dado risco particular.
- Em instituições financeiras, pode detectar os grupos de risco para a subscrição de um crédito.

# Regressão Logística

Tranformação Logit

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right)$$

$$g(x) = ln\left(\frac{\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}{1 - \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}\right) = ln\left(\frac{\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}{\frac{1}{1+e^{\beta_0+\beta_1 x}}}\right)$$

$$g(x) = ln\left(e^{\beta_0+\beta_1 x}\right) = \beta_0 + \beta_1 x$$