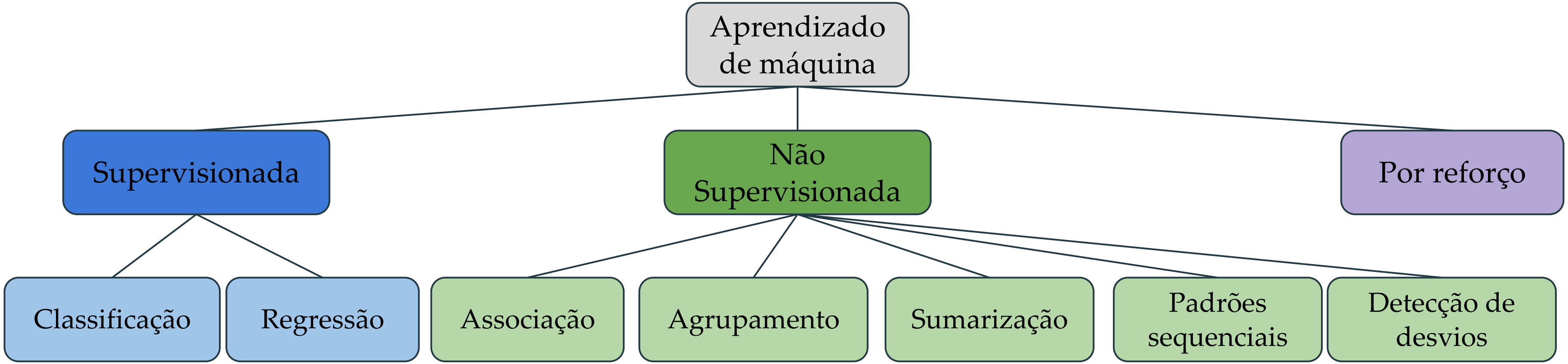
# Regressão Logística

Baseado no material produzido por Andrew Ng e adaptado por Alcione de Paiva Oliveira DPI/UFV

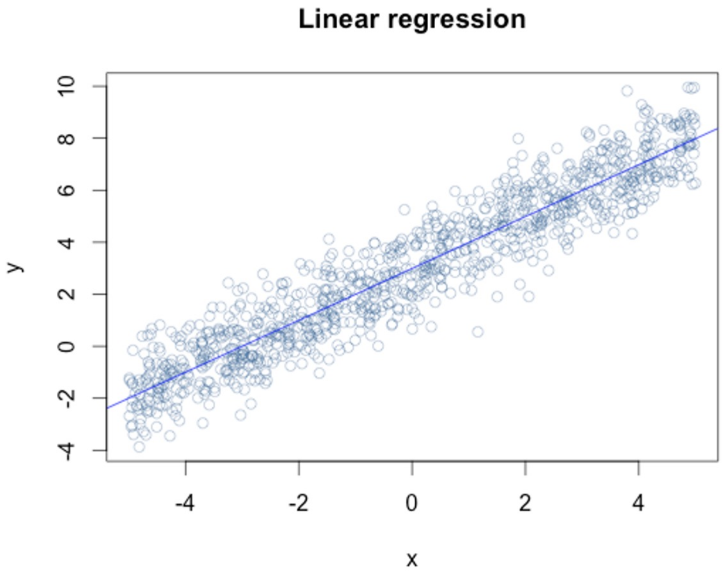
# Tipos de Aprendizado de Máquina



# Regressão Logística

Regressão linear e logística

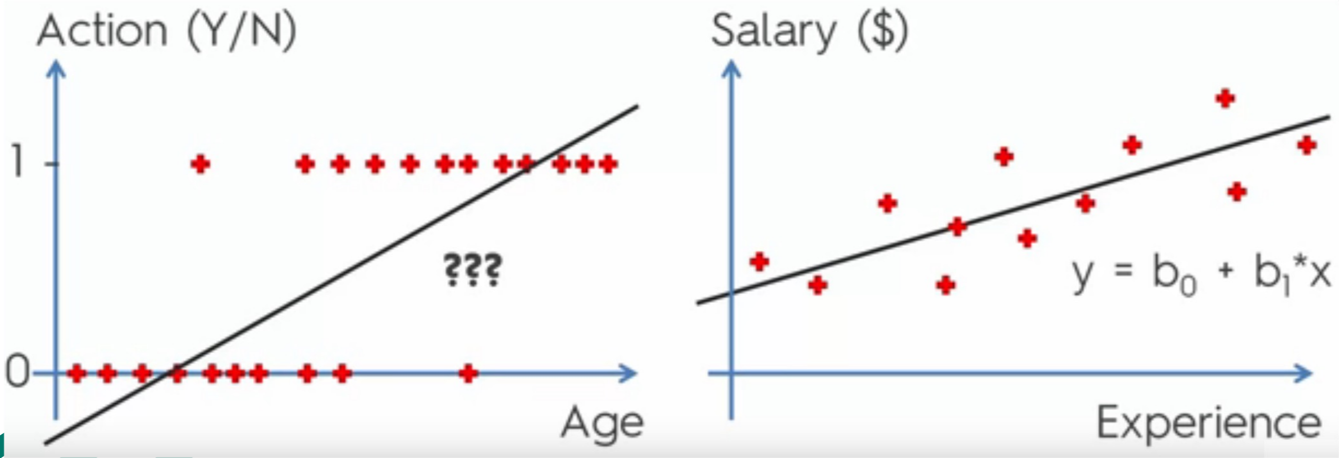
- Usa-se regressão linear quando se deseja prever um valor real a partir de uma função linear ajustada para valores de treinamento.



# Regressão Logística

Regressão linear e logística

- No entanto, quando se deseja obter a probabilidade de classificação a regressão logística é a adequada.



# Regressão Logística

**Regressão linear e logística**

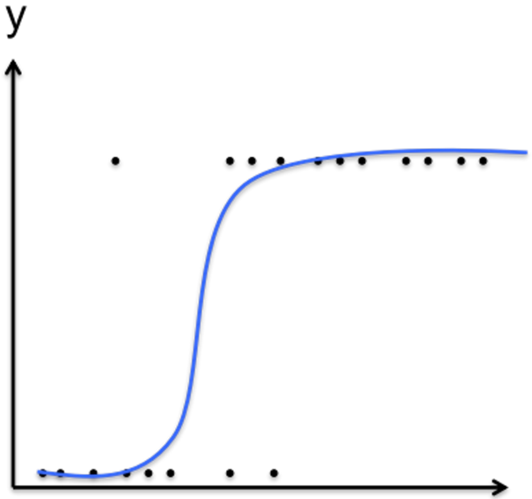
- No entanto, quando se deseja obter a probabilidade de classificação a regressão logística é a adequada.

Dado o salário, deve-se aprovar um empréstimo (S/N)

# Regressão Logística

**Regressão linear e logística**

- No entanto, quando se deseja obter a probabilidade de classificação a regressão logística é a adequada.



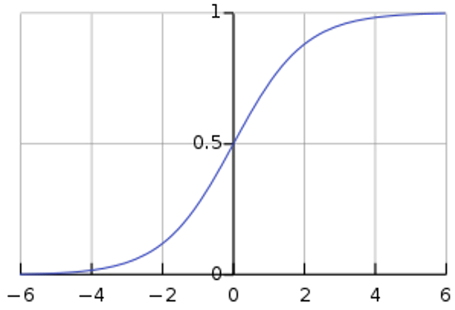
# Regressão Logística

**Regressão logística**

- A regressão logística usa a função logística, de forma a manter infinitos scores entre 0 e 1.

- Uma função logística ou uma curva logística é um formato de “S” comum (curva sigmóide), com equação :

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



# Regressão Logística

**Regressão logística**

- Então, suponha um vetor de entradas $\boldsymbol{x} = \{x_1, x_{2,=}, .., x_n\}$ e um vetor de pesos $\boldsymbol{w}= \{w_1, w_2, ..., w_n\}$ e uma constante $b$ (também denominado de bias), a função linear $z$ é definida como $z= \boldsymbol{w}^\mathsf{T}\boldsymbol{x} + b$.

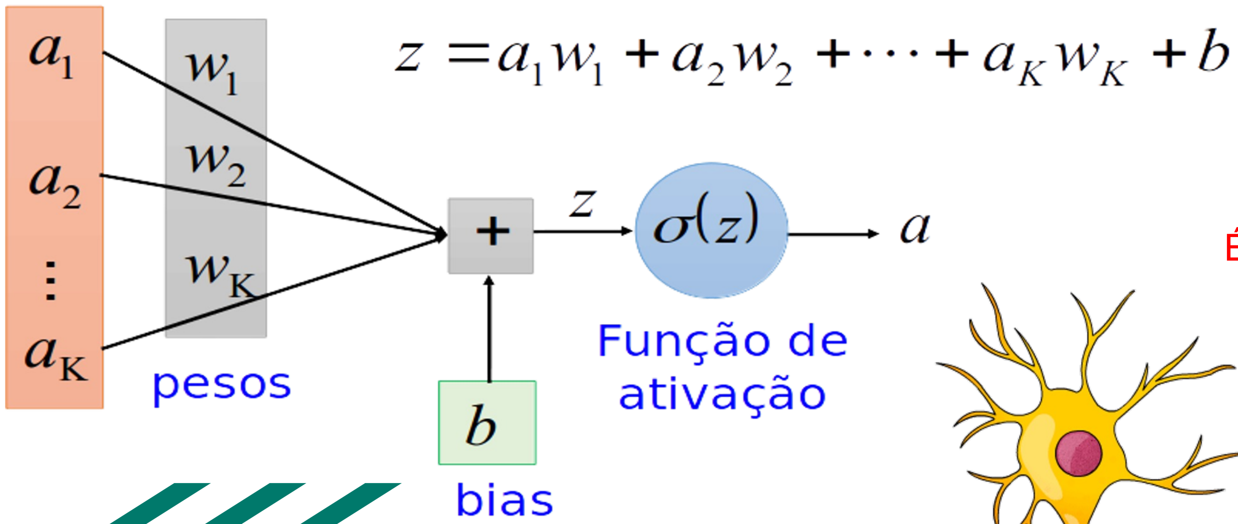
- Ao aplicarmos a função logística (função de ativação) σ à $z$ obtemos um valor entre 0 e 1:

$$\sigma\ (z) = \sigma\ (z= \boldsymbol{w}^\mathsf{T}\boldsymbol{x} + b)$$

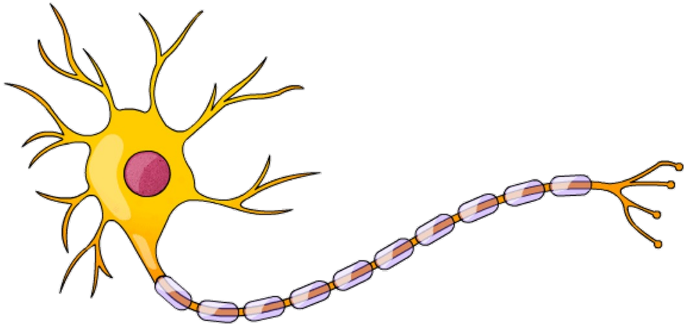
# Regressão Logística

**Regressão logística**

$$z = a_1 w_1 + a_2 w_2 + \cdots + a_K w_K + b$$



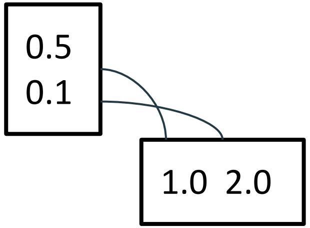

É um neurônio artificial



# Regressão Logística

**Regressão logística: Exemplo**

- Vetor com valor de salário e valor pedido de empréstimo (em unidade de milhar) $\boldsymbol{x} = \{1{,}2\}$ e um vetor de pesos $\boldsymbol{w} = \{0.5{,}0.1\}$ e uma constante $b$=0.1, então $z$, definida como $z = \boldsymbol{w}^{\mathsf{T}}\boldsymbol{x} + b$ é




$$1\times 0.5 + 2\times 0.1 + 0.1 = 0.8$$

- Ao aplicarmos a função logística (função de ativação) σ à $z$ obtemos um valor entre 0 e 1:

$$\sigma(z) = \sigma(z = \boldsymbol{w}^{\mathsf{T}}\boldsymbol{x} + b) = 1/(1+e^{-0.8}) = 0.69$$

# Regressão Logística

**Regressão logística**

- Mas como saber se o valor emitido está correto?
- O valor de 0.69 (69%) para emprestar para uma pessoa que ganha mil Reais e pediu dois mil empréstimos está correto?
- Como é feito o ajuste (aprendizado) da regressão logística?

# Regressão Logística

**Regressão logística**

- O aprendizado é feito por meio do ajuste dos pesos.
- O ajuste é feito por meio da correção em etapas, dos pesos, a partir  erro apresentado na saída de exemplos, cuja saída é conhecida (aprendizado supervisionado)

# Regressão Logística

**Função de perda (loss function)**

$\hat{y}$ = σ ($z$= $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{x}$ +$b$) onde σ ($z$) = $1/(1+e^{-z})$

Dado um conjunto $C = \{(\boldsymbol{x}^{(1)},y^{(1)}),\ldots,(\boldsymbol{x}^{(m)},y^{(m)})\}$ queremos que $\hat{y}^{(i)} \approx y^{(i)}$.

A loss function $L(\hat{y},y)$ indica o quanto se errou.

Existem várias funções de perda:

Por que elevar ao quadrado?

$L(\hat{y},y) = 1/\mathrm{n} \sum_{i \in C}( \hat{y}^{(i)} - y^{(i)})^2$ (Erro Médio Quadrático - MSE)

*Mas não é uma boa lf

# Regressão Logística

**Função de perda (loss function)**

$\hat{y}$ = σ ($z$= $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{x}$ +$b$) onde σ ($z$) = $1/(1+e^{-z})$

Dado um conjunto $C= \{(\boldsymbol{x}^{(1)},y^{(1)}),\ldots,(\boldsymbol{x}^{(m)},y^{(m)})\}$ queremos que $\hat{y}^{(i)} \approx y^{(i)}$.

A loss function $L(\hat{y},y)$ indica o quanto se errou.

Existem várias funções de perda:

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

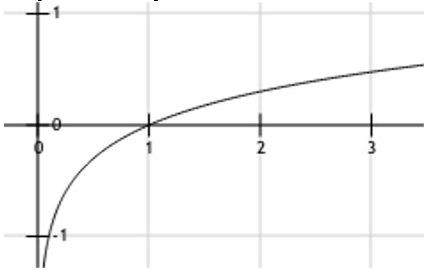
# Regressão Logística

**Função de perda (loss function)**

$\hat{y}$ = $\sigma$ ($z$= $\mathbf{w}^T\mathbf{x}$ +$b$) onde $\sigma$ ($z$) = $1/(1+e^{-z})$

Dado um conjunto C= $\{(\mathbf{x}^{(1)},y^{(1)}),\ldots,(\mathbf{x}^{(m)},y^{(m)})\}$ queremos que $\hat{y}^{(i)} \approx y^{(i)}$.

A loss function $L(\hat{y},y)$ indica o quanto se errou.




Existem várias funções de perda:

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

Se y = 1 então $\mathscr{L}$ = - $\log \hat{y}$ e para minimizar a perda $\hat{y}$ precisa ser grande
Se y = 0 então $\mathscr{L}$ = - $\log(1-\hat{y})$ e para minimizar a perda $\hat{y}$ precisa ser pequeno

# Regressão Logística

**Função de perda (loss function)**

$\hat{y} = \sigma\,(z = \mathbf{w}^{\mathsf{T}}\mathbf{x} + b)$ onde $\sigma\,(z) = 1/(1+e^{-z})$

Dado um conjunto $C = \{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(m)}, y^{(m)})\}$ queremos que $\hat{y}^{(i)} \approx y^{(i)}$.

A loss function $L(\hat{y}, y)$ indica o quanto se errou.

Existem várias funções de perda:

$$J(\mathbf{w}, b) = 1/m \ \sum_{i \in C}(\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -1/m \ \sum_{i \in C}[y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)})]$$

# Regressão Logística

Regressão logística (**Gradiente descendente**)

- **Uma vez detectado o erro como ajustar  a função para minimizar o erro?**

Ajustando os pesos $\boldsymbol{w}$

- Podemos ajustar os pesos alterando no sentido inverso da inclinação do hiperplano do erro. (Gradient Descent)
- Obtemos a inclinação derivando (ou obtendo a *derivada direcional* no caso de vetor) a função de perda.

# Regressão Logística

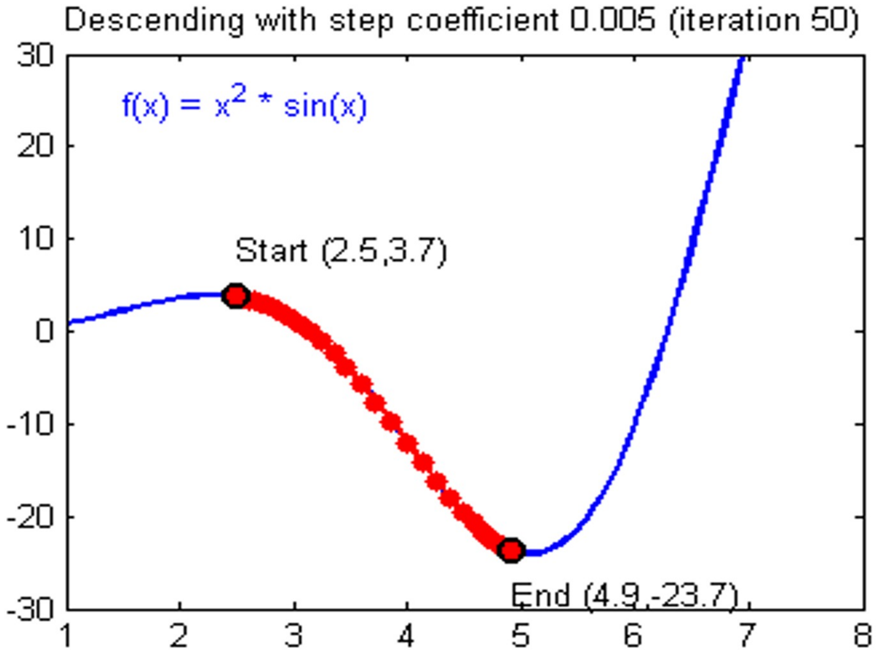
Regressão logística (**Gradiente descendente**)

$f'(x) = 2x\sin(x) + x^2\cos(x)$

*No ponto* (2.8,2.6) f'(x) = -5.511

Então ajustamos no sentido inverso:

x= 2.8 -(0.005 × -5.511) = 2.82



# Regressão Logística

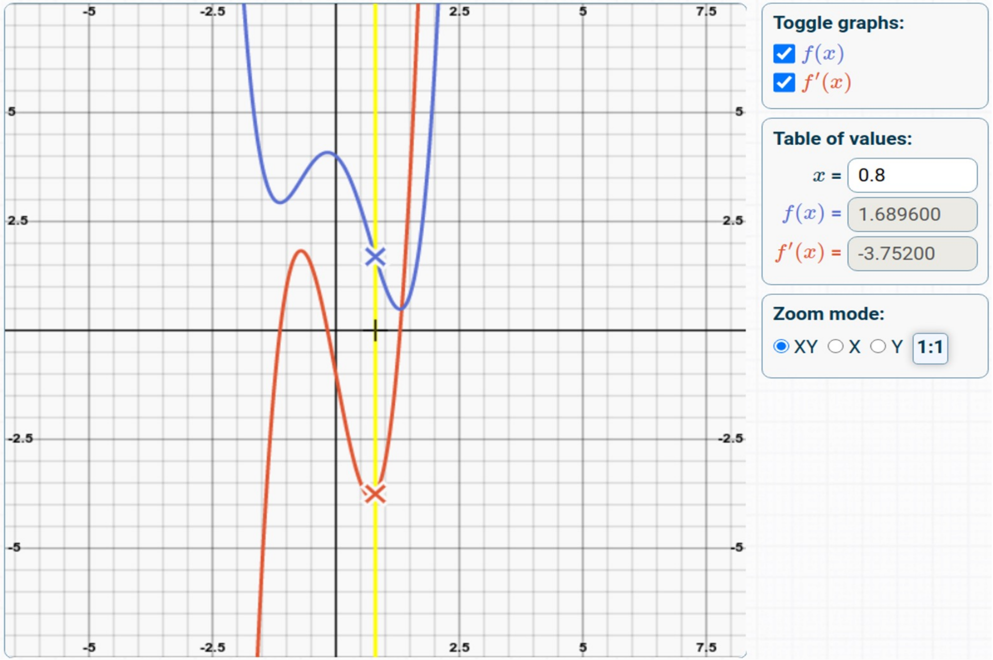
Regressão logística (**Gradiente descendente**) **Outro Exemplo**

$$f(x) = x^4 - 3x^2 - x + 4$$

$$f'(x) = 4x^3 - 6x - 1$$

Se ajustarmos ajustamos com alpha muito grande, iremos passar o ponto mínimo:

x= 0.8 -(0.3 × -3.75) = 1.9




https://www.derivative-calculator.net/

# Regressão Logística

Regressão logística (**Gradiente descendente**)

- No caso de vetor e no âmbito do cálculo vetorial a inclinação é dada pelo gradiente da superfície de erro.

- O Gradiente de uma função $f(\mathbf{x})$: $\mathbb{R}^n \rightarrow \mathbb{R}$ é um vetor da seguinte forma:

$$\nabla f(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)$$

# Regressão Logística

Regressão logística (**Gradiente descendente**)

- O que desejamos é onde no ponto $a$ o gradiente é o mais inclinado.
- Ou seja dado um vetor unitário $\boldsymbol{u}$ que parte do ponto qual é a inclinação de $\boldsymbol{u}$. (derivativa direcional)

$$D_{\mathbf{u}}f(\mathbf{a}) = \lim_{h \to 0} (f(\boldsymbol{a}+h\boldsymbol{u}) - f(\boldsymbol{a}))/h$$

# Regressão Logística

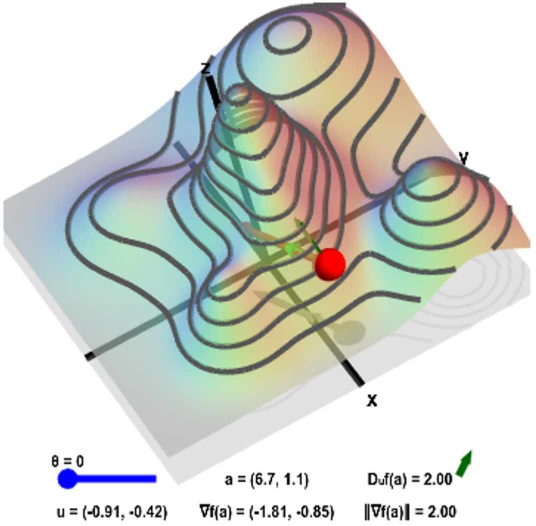
Regressão logística (**Gradiente descendente**)

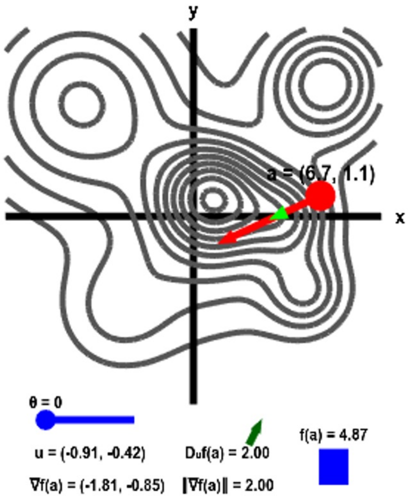
- $D_{\mathbf{u}}f(\mathbf{a})$ é máximo (*m* maior inclinação) é dada pela fórmula:

$$D_{\mathbf{m}}f(\mathbf{a}) = \nabla f(\boldsymbol{a}) / ||\nabla f(\boldsymbol{a})||$$

# Regressão Logística

Regressão logística (Gradiente descendente)




https://mathinsight.org/applet/gradient_directional_derivative_mountain

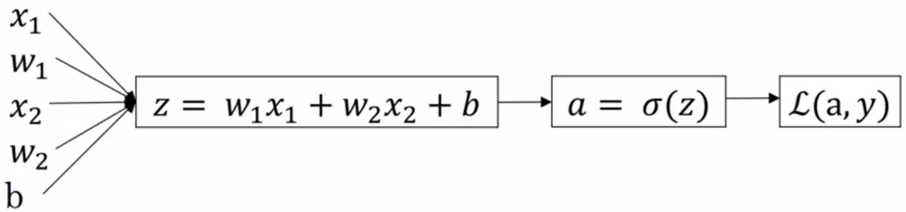
# Regressão Logística

Regressão logística (Gradiente descendente)

$$z = w^T x + b$$

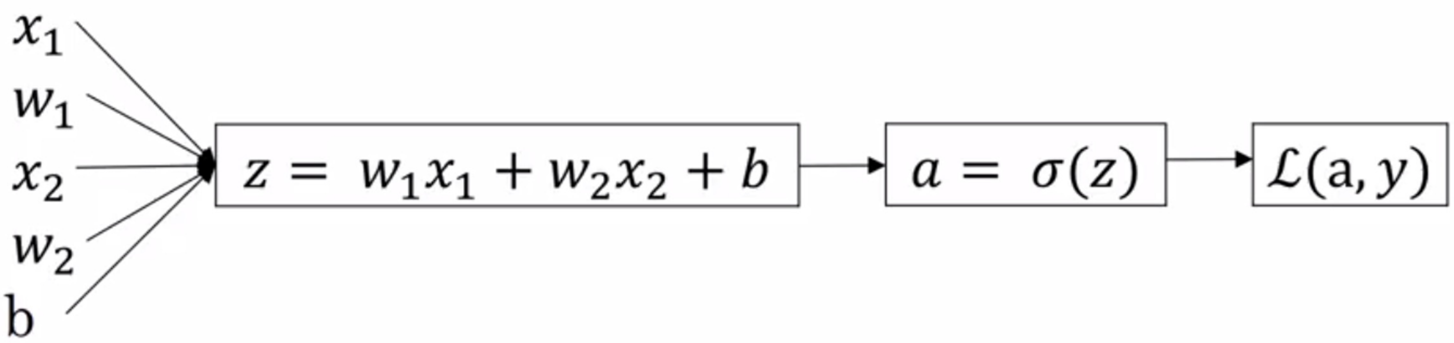
$$\hat{y} = a = \sigma(z)$$

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



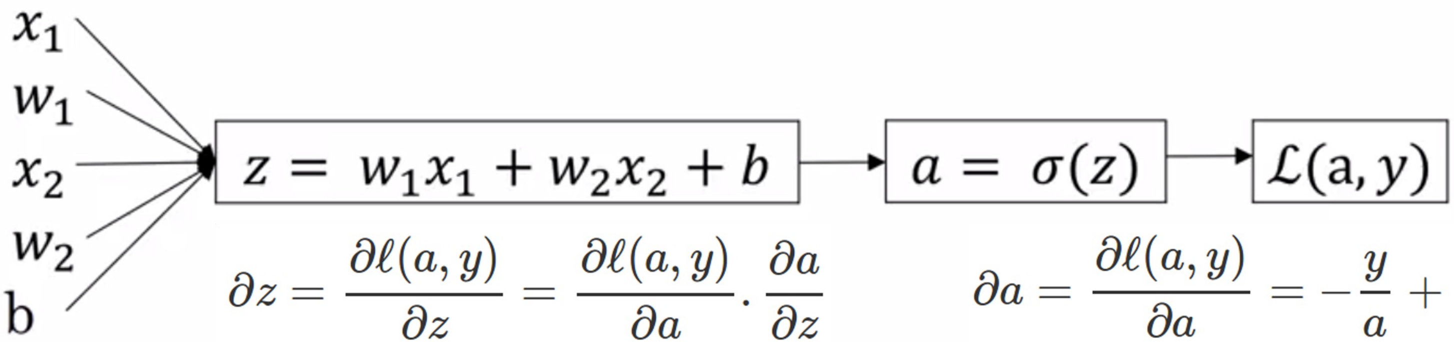
# Regressão Logística

Regressão logística (Gradiente descendente)



# Regressão Logística

Regressão logística (Gradiente descendente)




$$\partial z = \frac{\partial \ell(a,y)}{\partial z} = \frac{\partial \ell(a,y)}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$$\partial a = \frac{\partial \ell(a,y)}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$$

Onde: $\frac{\partial a}{\partial z} = a(1-a)$

Logo: $\partial z = a - y$

# Regressão Logística

Regressão logística (Gradiente descendente)

**Regra da cadeia**

Seja $F = f \circ g$, ou de forma equivalente, $F(x) = f(g(x))$, então

$$F'(x) = f'(g(x))g'(x).$$

ou

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

# Regressão Logística

Regressão logística (Gradiente descendente)

**Regra da cadeia**

Exemplo: $y = e^{\sin x^2}$.

$$y = f(u) = e^u,$$
$$u = g(v) = \sin v,$$
$$v = h(x) = x^2.$$

$$\frac{dy}{du} = f'(u) = e^u,$$
$$\frac{du}{dv} = g'(v) = \cos v,$$
$$\frac{dv}{dx} = h'(x) = 2x.$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dv} \cdot \frac{dv}{dx}.$$

$$\frac{dy}{dx} = e^{\sin x^2} \cdot \cos x^2 \cdot 2x.$$

# Regressão Logística

Regressão logística (Gradiente descendente)

$x_1, w_1, x_2, w_2, b \rightarrow \boxed{z = w_1 x_1 + w_2 x_2 + b} \rightarrow \boxed{a = \sigma(z)} \rightarrow \boxed{\mathcal{L}(a, y)}$

$$\partial z = \frac{\partial \ell(a,y)}{\partial z} = \frac{\partial \ell(a,y)}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$$\partial a = \frac{\partial \ell(a,y)}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$$

Onde: $\frac{\partial a}{\partial z} = a(1-a)$

Logo: $\partial z = a - y$

$$\frac{\partial z}{\partial w_1} = dw_1 = x_1 dz \qquad dw_2 = x_2 dz \qquad db = dz$$

# Regressão Logística

Regressão logística (Gradiente descendente)

$$\begin{array}{l} x_1 \\ w_1 \\ x_2 \\ w_2 \\ b \end{array} \longrightarrow \boxed{z = w_1x_1 + w_2x_2 + b} \longrightarrow \boxed{a = \sigma(z)} \longrightarrow \boxed{\mathcal{L}(a, y)}$$

$$\partial z = \frac{\partial \ell(a,y)}{\partial z} = \frac{\partial \ell(a,y)}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$$\partial a = \frac{\partial \ell(a,y)}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$$

Onde: $\frac{\partial a}{\partial z} = a(1-a)$ Logo: $\partial z = a - y$

$$w_1 := w_1 - \alpha\, dw_1$$
$$w_2 := w_2 - \alpha\, dw_2$$
$$b := b - \alpha\, db.$$

$$\frac{\partial z}{\partial w_1} = dw_1 = x_1 dz \quad dw_2 = x_2 dz \quad db = dz$$

# Regressão Logística

**FIM**