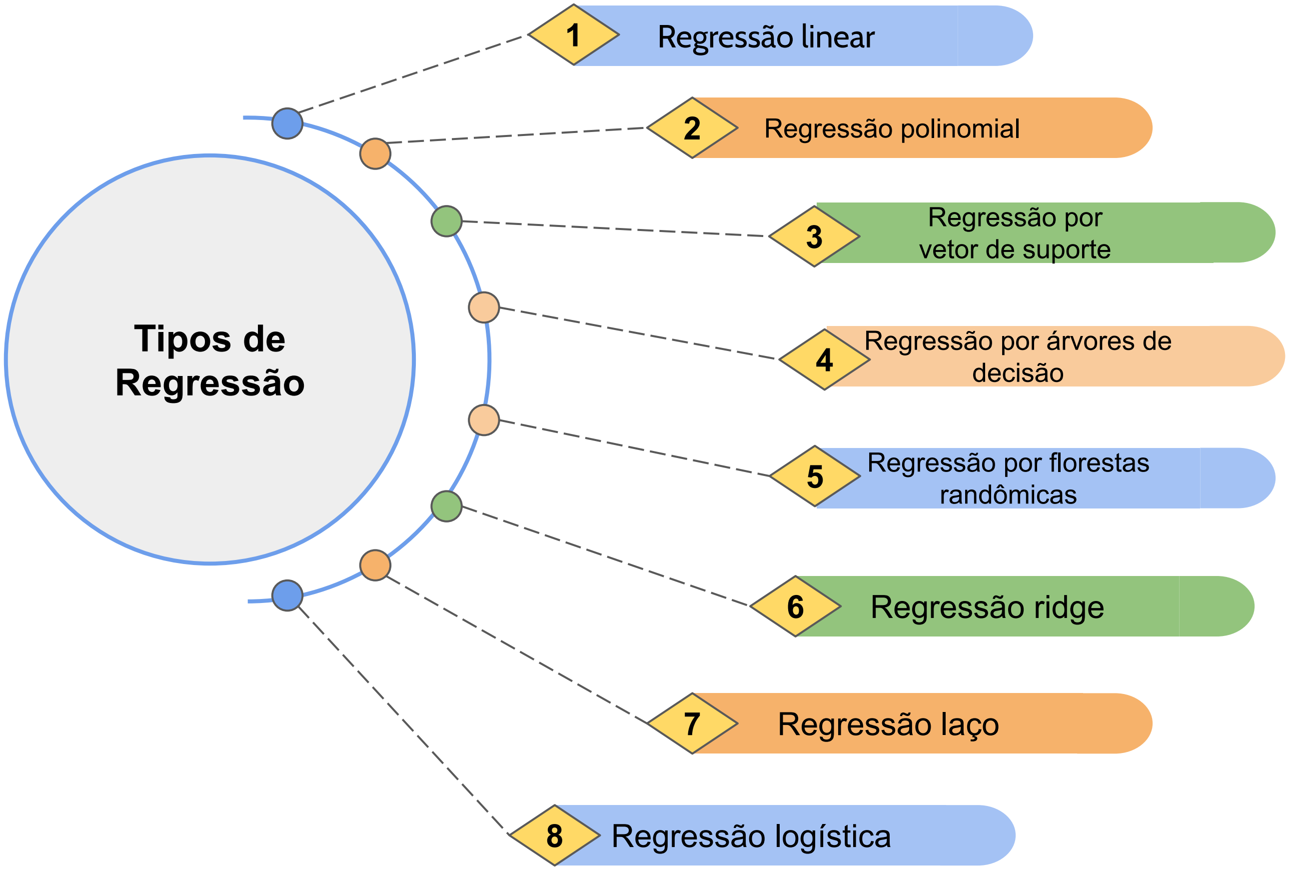
Aula 2b
Regressão Linear

Adaptado dos slides do Prof Samuel Botter

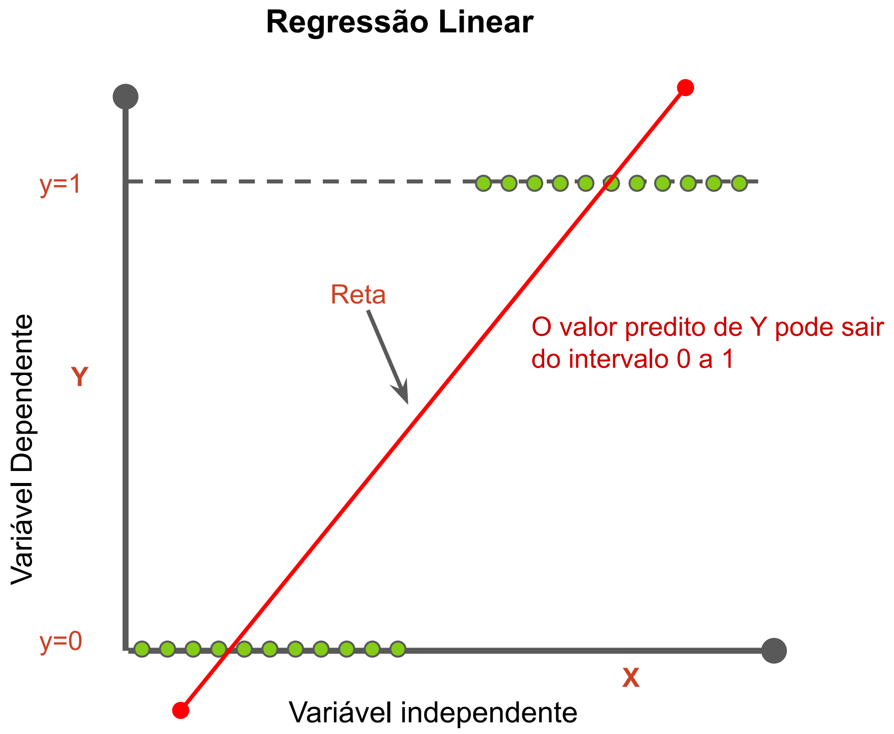
O que é uma regressão?

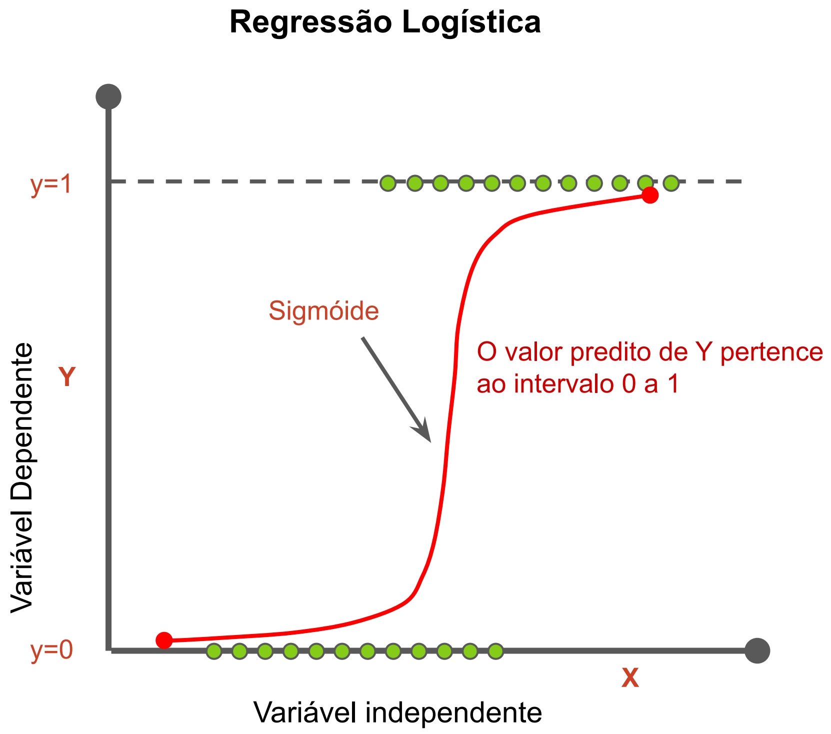
- É uma abordagem para modelar o relacionamento entre variáveis numéricas independentes (explicativas) e dependentes, ajustando um modelo (uma curva qualquer) para as observações de um conjunto de treinamento.
- Este modelo pode então ser usado para prever aquelas variáveis numéricas dependentes para novas observações ainda não observadas das variáveis explicativas
- Quando esta curva é uma reta, diz-se que a regressão é **linear**

Tipos de regressão



Regressão Linear X Logística





Variáveis dependentes e independentes

	"causas" variáveis independentes		"efeito" variável independente
Idade	**Titulação**	**Experiência**	**Salário Anual**
21	Graduação	1	35.000,00
25	Especialização	5	80.000,00
35	Doutorado	10	120.000,00
...	...	...	...

Conjunto de Treinamento

Uma variável independente

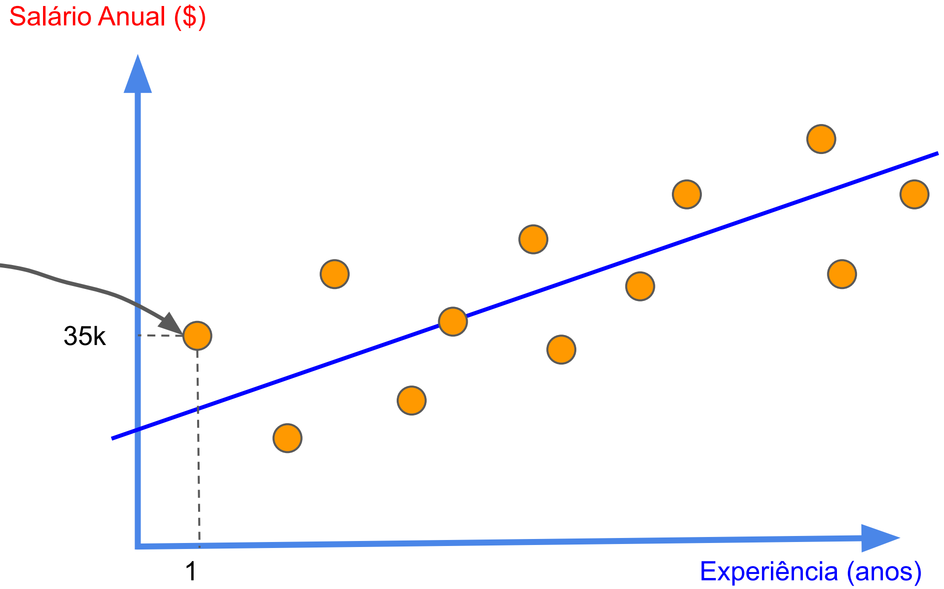
$$\hat{y} = b_0 + b_1 \cdot x$$

Modelo Linear

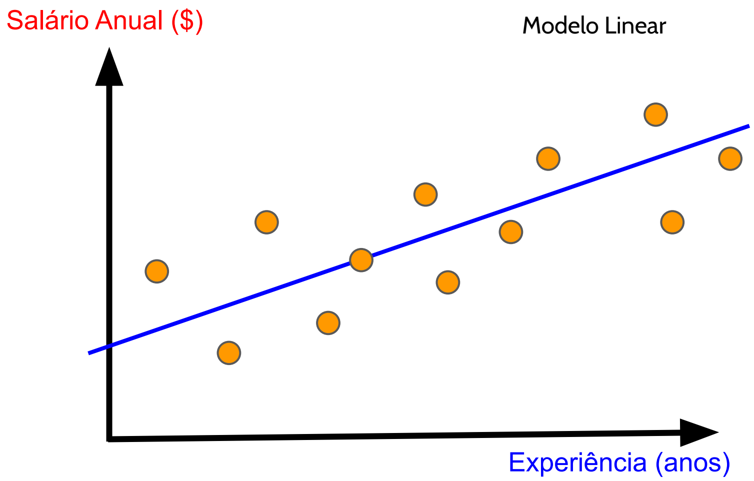
variáveis independentes | variável dependente

Experiência	Salário Anual
1	35.000,00
5	80.000,00
10	120.000,00
...	...

Conjunto de Treinamento



Regressão Linear Simples



$$\underbrace{h(x_1)}_{hipótese} = \overbrace{\hat{y}}^{\substack{variável \\ dependente}} = \overbrace{b_0}^{intercepto} + \underbrace{b_1}_{inclinação} \cdot \overbrace{x}^{\substack{variável \\ independente}}$$

$$salário = b_0 + b_1 \cdot experiência$$

Equação da Reta

$$(y - y_0) = \mu \cdot (x - x_0)$$

$$\text{simplificando para } \mathrm{x}_0 = 0$$

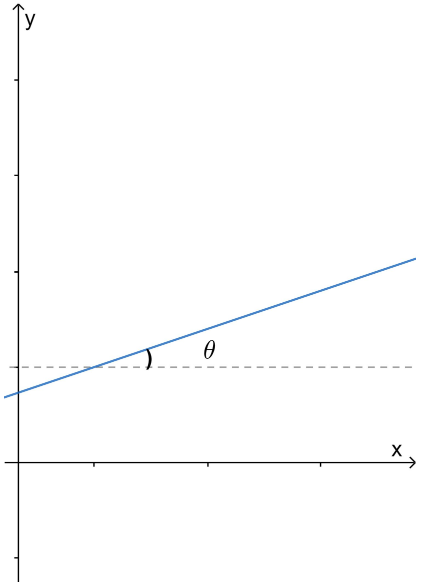
$$(y - y_0) = \mu \cdot x$$

$$y = \overbrace{y_0}^{intercepto} + \underbrace{\mu}_{inclinação} \cdot x$$

$$h(x) = \hat{y} = \overbrace{b_0}^{intercepto} + \underbrace{b_1}_{inclinação} \cdot x$$

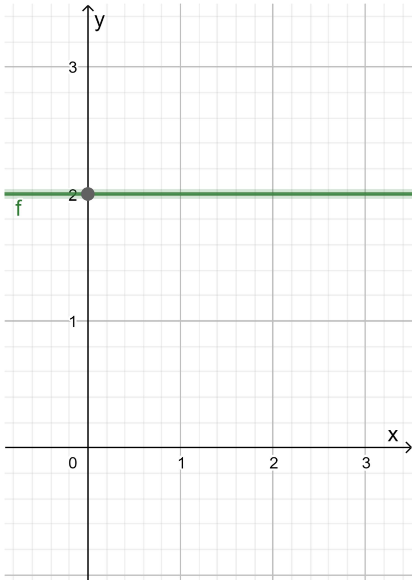
$$\mu = b_1 = \tan(\theta)$$

$$\theta = \arctan(\mu)$$

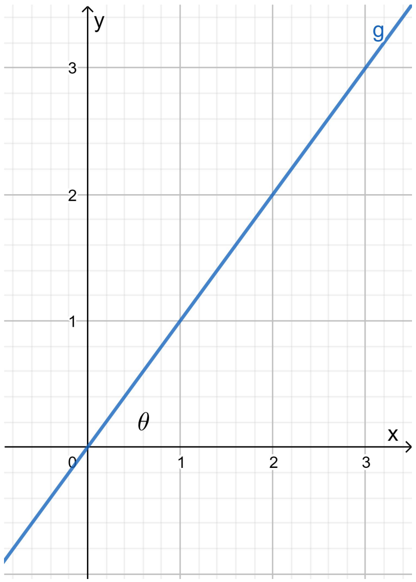


Equação da Reta : exemplos

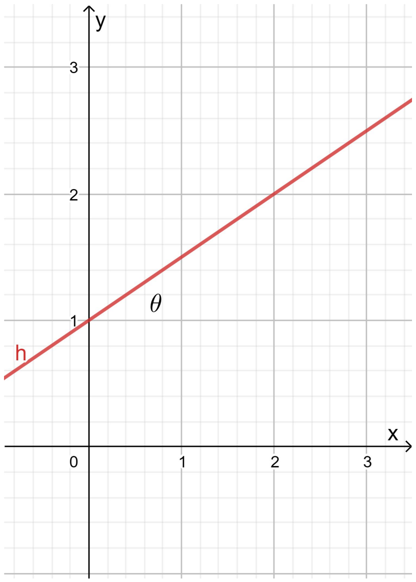
$$h(x) = \hat{y} = \overbrace{b_0}^{intercepto} + \underbrace{b_1}_{inclinação} \cdot x$$



$b_0 = 2.0$
$b_1 = 0.0$

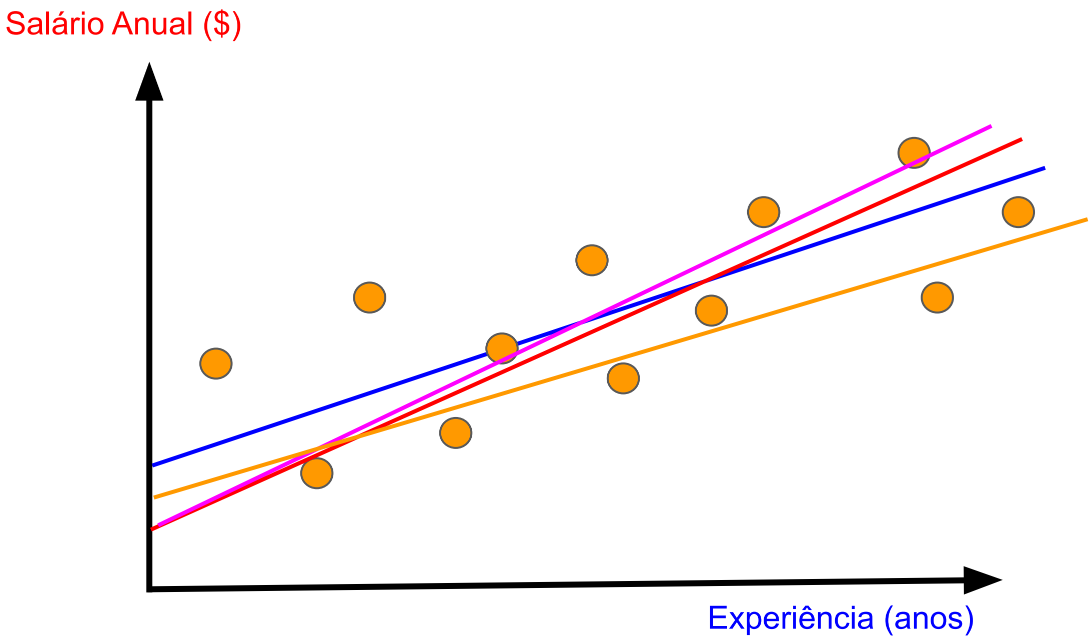


$b_0 = 0.0$
$b_1 = 1.0$
$(\theta = 45^\circ)$



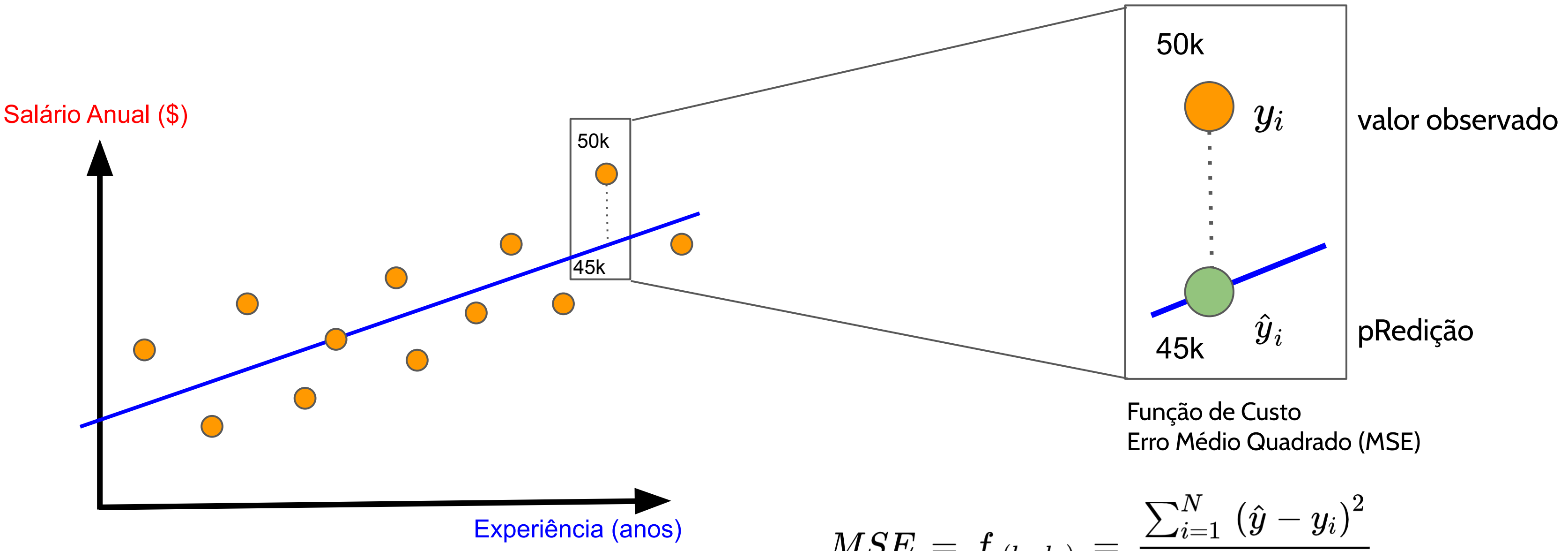
$b_0 = 1.0$
$b_1 = 0.5$
$(\theta \approx 27^\circ)$

Qual a melhor reta ?



Erro do modelo linear

Objetivo: achar os parâmetros b_0 e b_1 que minimizam a função de custo para o conjunto de treinamento



Função de Custo
Erro Médio Quadrado (MSE)

$$MSE = f_{c(b_0,\, b_1)} = \frac{\sum_{i=1}^{N} (\hat{y} - y_i)^2}{N}$$

(Mean SquaRed Error)

A reta de menor valor da função de custo

$$\text{Modelo linear (hipótese)}$$
$$h(x) \; = \; \hat{y} \; = \; b_0 + b_1 \cdot x$$

$$\text{Modelo linear Simplificado}$$
$$h(x) \; = \; \hat{y} \; = \; b_1 \cdot x$$
$$\text{simplificando } b_0 = 0$$

$$\text{Função de Custo}$$
$$fc(b_o, b_1) \; = \; MSE \; = \; \frac{\sum_{i=1}^{N} \left(\hat{y} - y_i\right)^2}{N}$$

$$\text{Função de Custo}$$
$$fc(b_1) \; = \; MSE \; = \; \frac{\sum_{i=1}^{N} \left(\hat{y} - y_i\right)^2}{N}$$

$$Objetivo$$
$$\min_{b_0, b_1} \left(fc\left(b_0, b_1\right)\right)$$

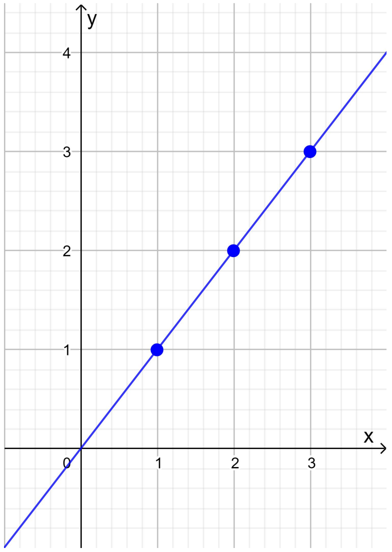
$$Objetivo$$
$$\min_{b_1} \left(fc\left(b_1\right)\right)$$

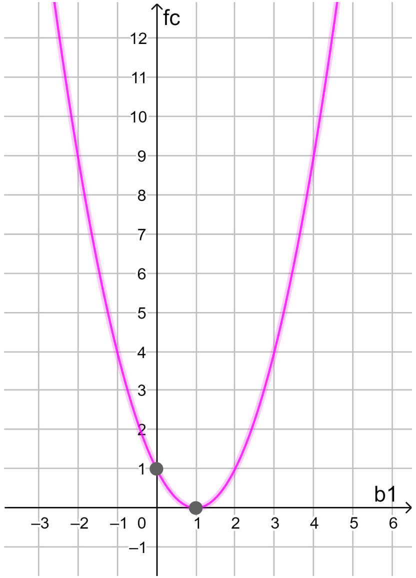
Exemplo de função de custo

$$h(x) = \hat{y} = b_1 \cdot x$$

$$b_1 = 1$$

$$fc(b_1) = MSE = \frac{\sum_{i=1}^{N} \left(\overbrace{b_1 \cdot x_i}^{\hat{y}} - y_i \right)^2}{N}$$





Resumo da ideia de minimização

É um processo de otimização.
A função de custo (que deve ser minimizada) é o erro médio quadrático.

Modelo linear (hipótese)

$$h(x) = \hat{y} = b_0 + b_1 \cdot x$$

Função de Custo

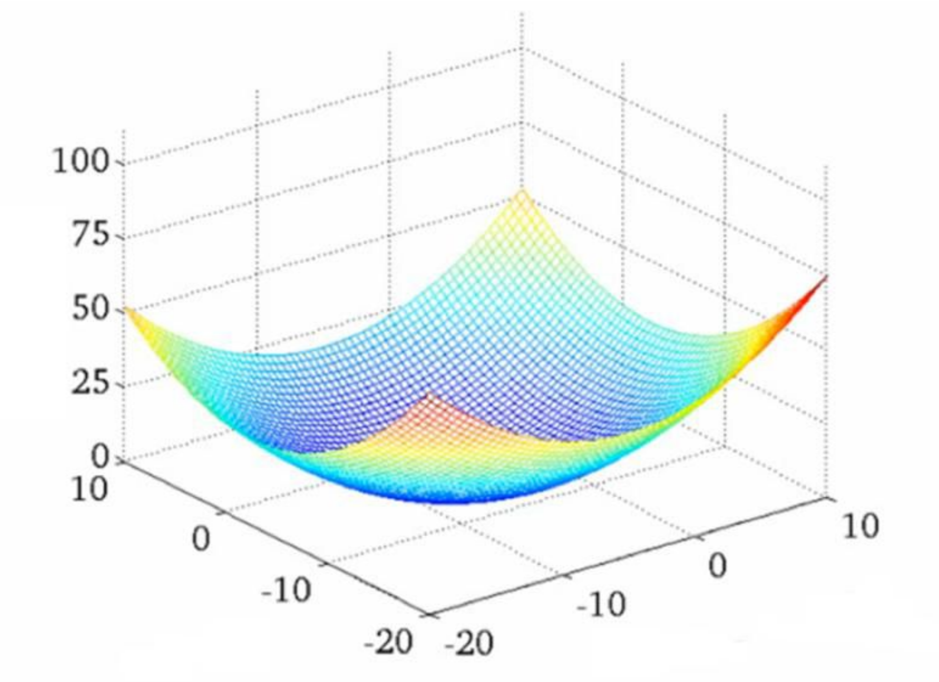
$$fc(b_o, b_1) = MSE = \frac{\sum_{i=1}^{N} (\hat{y} - y_i)^2}{N}$$

Objetivo

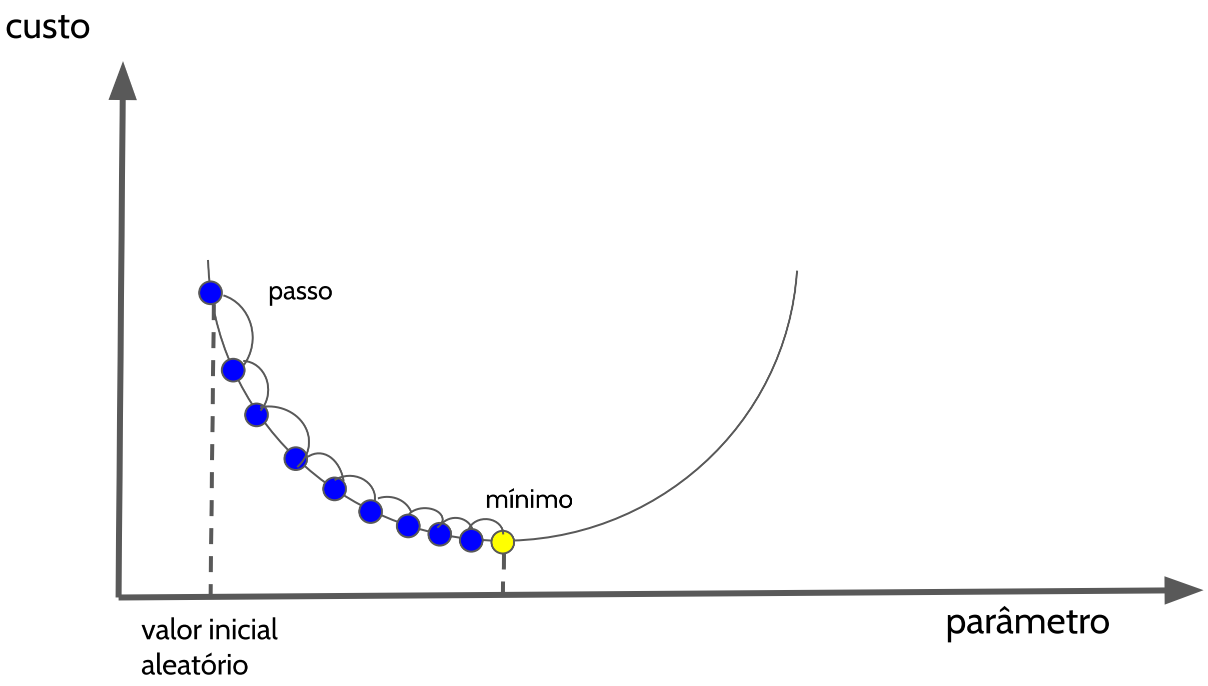
$$\min_{b_0, b_1} (fc(b_0, b_1))$$

Função de custo

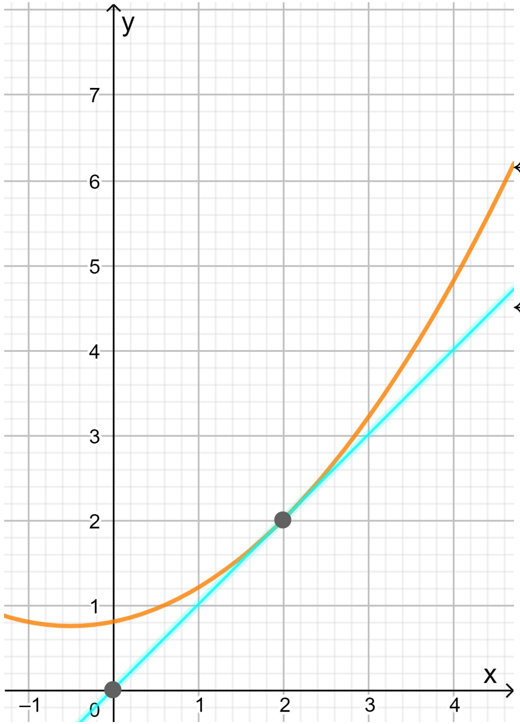
$$fc\,(b_0,\,b_1) = MSE = \frac{\sum_{i=1}^{N}\left(\overbrace{b_0 + b_1 \cdot x_i}^{\hat{y}} - y_i\right)^2}{N}$$



Gradiente Descendente



Derivada de uma função



$$\leftarrow \; y = f(x)$$

$$\leftarrow \; \text{inclinação} = \frac{\mathrm{d}y}{\mathrm{d}x} =$$

$$= f'(x_1) =$$

$$= \dot{y} = y'$$

$$x_1 = 2$$

A derivada de uma função $y = f(x)$ em um dado ponto x_1 é a inclinação da **reta tangente à curva** no ponto x_1 ,

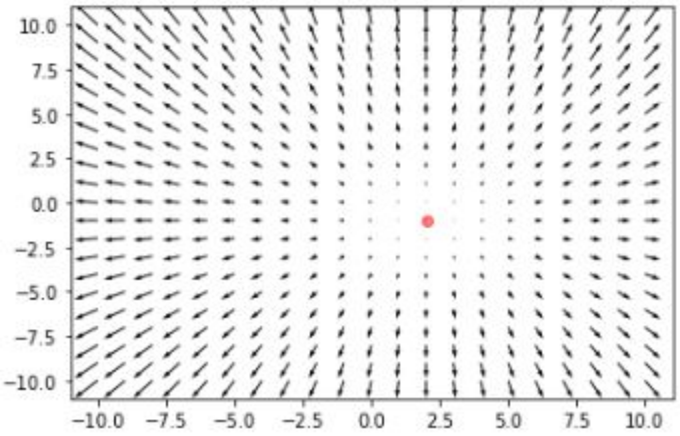
isto é, é a **tangente do ângulo entre a reta tangente** à curva no ponto x_1 e o eixo x

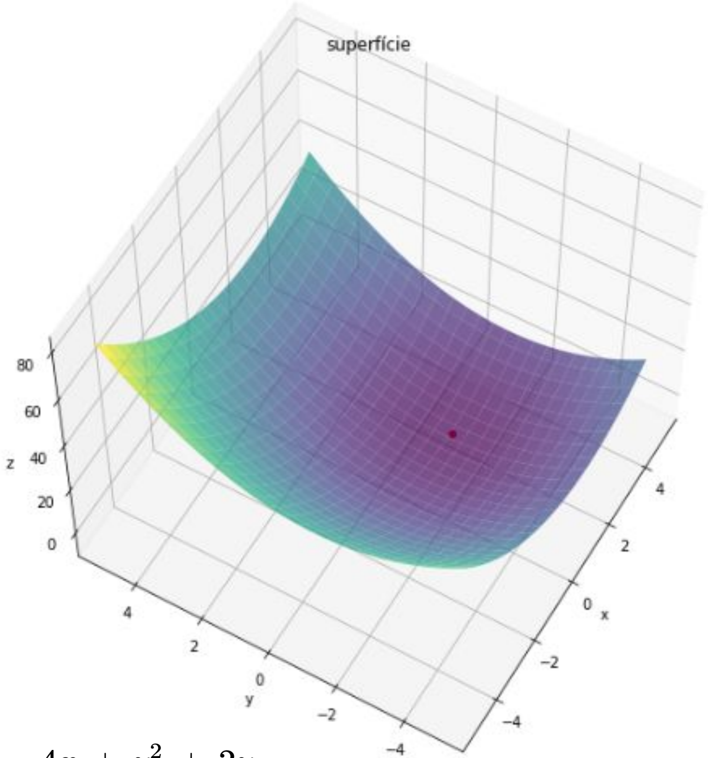
Vetor Gradiente

- O **vetor gradiente** (ou apenas gradiente), é um vetor que aponta para a direção de maior variação da função em um determinado ponto

- É formado pelas derivadas parciais de cada variável da função

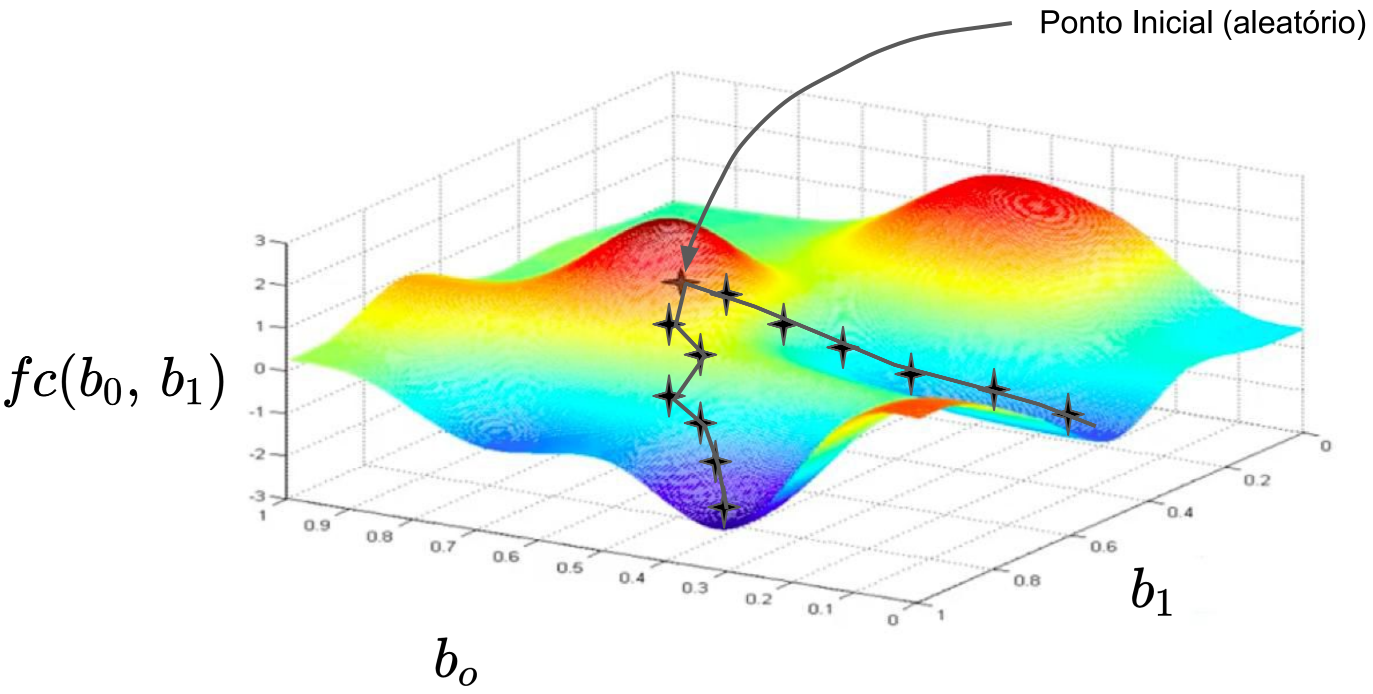
O campo gradiente $\langle 2x-4,\, 2y+2 \rangle$





da função $f = x^2 - 4x + y^2 + 2y$

Gradiente Descendente



Gradiente Descendente: algoritmo

Dada uma função de custo $fc(b_0, b_1)$

O objetivo é $\min\limits_{b_0, b_1} (fc(b_0, b_1))$

1. inicie os parâmetros $b_0,\ b_1$ (por exemplo, $b_0 = 0,\ b_1 = 0$);
2. varie b_0, b_1 de modo a reduzir $fc(b_0, b_1)$

   enquanto o valor de fc for menor que o anterior
3. quando o valor de fc não se reduzir mais, chegou-se a um mínimo (local ou global)

Gradiente Descendente: algoritmo

taxa de aprendizado

Repita

$$temp_0 = b_0 - \alpha \frac{\partial}{\partial b_0} fc(b_0, b_1)$$

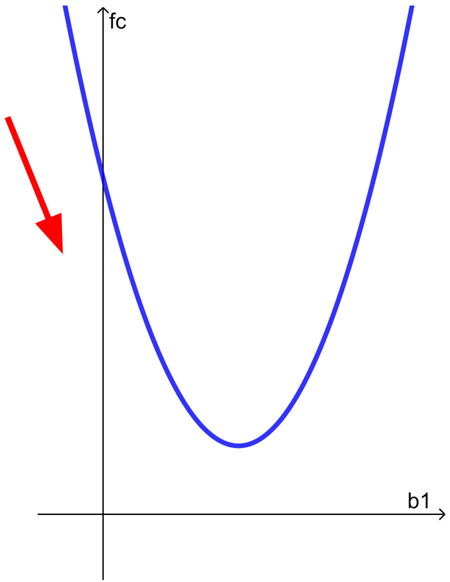
$$temp_1 = b_1 - \alpha \frac{\partial}{\partial b_1} fc(b_0, b_1)$$

$$b_0 = temp_0$$

$$b_1 = temp_1$$

até convergir

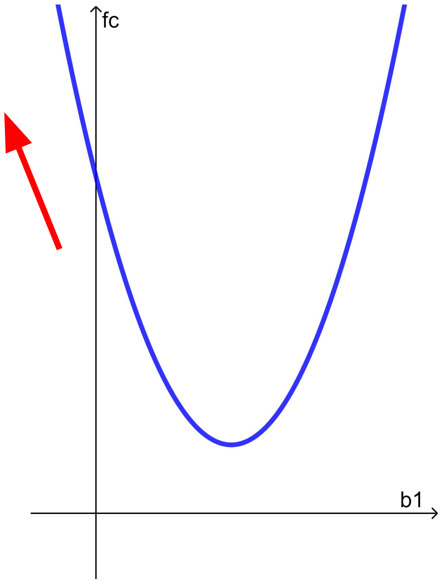
A taxa de aprendizado



$$b_1 = b_1 - \alpha \frac{\partial}{\partial b_1} fc(b_1)$$

$$\uparrow$$

$$\geqslant 0$$

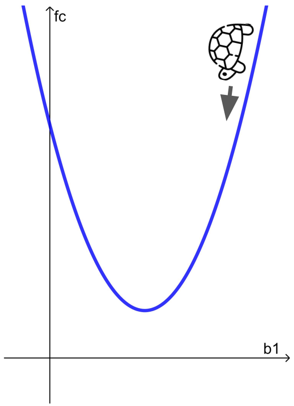


$$b_1 = b_1 - \alpha \frac{\partial}{\partial b_1} fc(b_1)$$

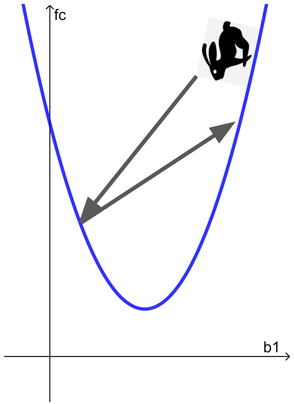
$$\uparrow$$

$$\leq 0$$

A taxa de aprendizado



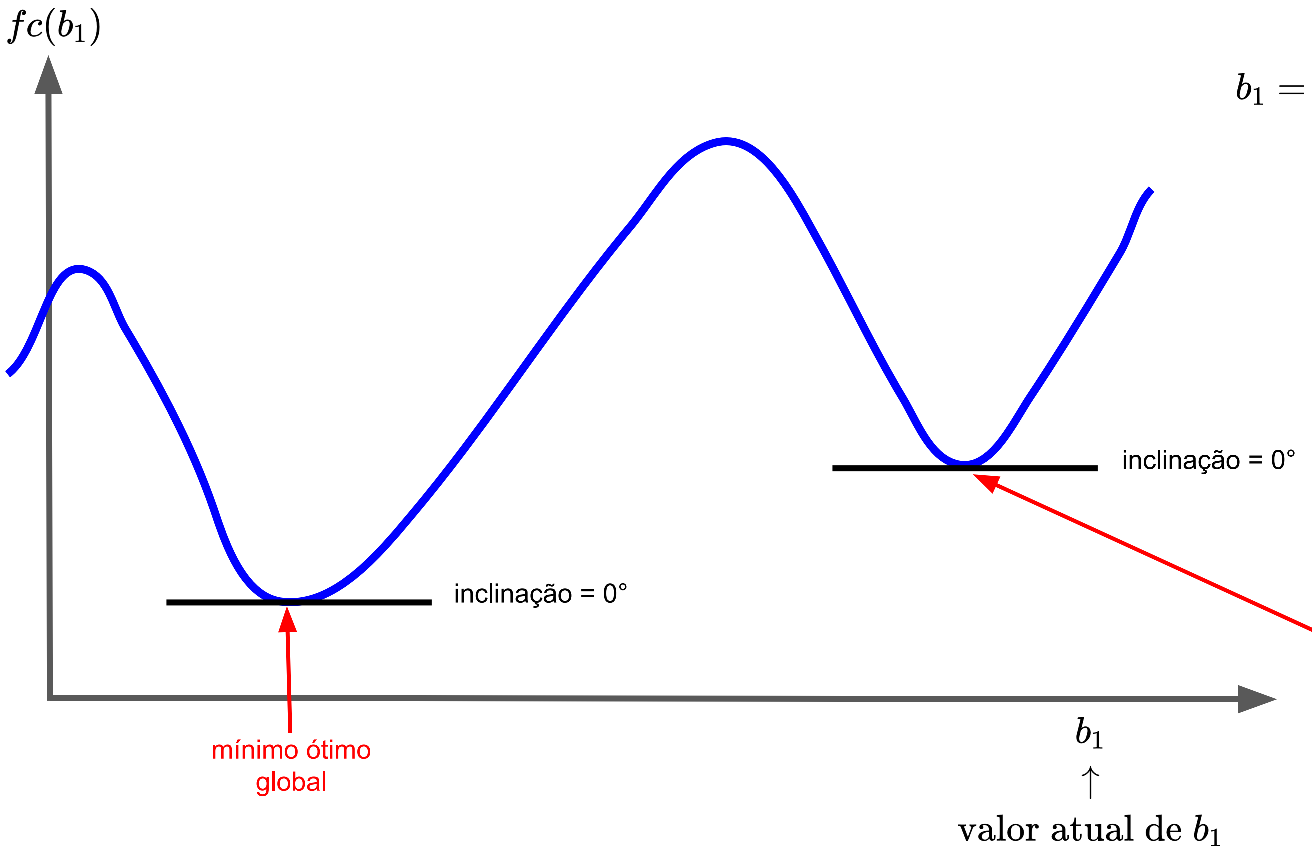
$$b_1 = b_1 - \alpha \frac{\partial}{\partial b_1} fc(b_1)$$



Se α é muito pequeno, o gradiente descendente pode ser **lento**

Se α é muito grande, o gradiente descendente pode sempre **passar pelo mínimo** em ambos os sentidos. A convergência pode falhar

Mínimos locais e globais



$$b_1 = b_1 - \alpha \frac{\partial}{\partial b_1} fc(b_1)$$

$$\tan(0) = 0$$

$$\frac{\partial}{\partial b_1} fc(b_1) = 0$$

Gradiente Descendente para Regressão Linear

Embora o algoritmo de gradiente descendente possa convergir para um <u>mínimo local</u>, isto não será problema para a regressão linear, uma vez que estamos utilizando uma função de custo que é uma **função convexa**