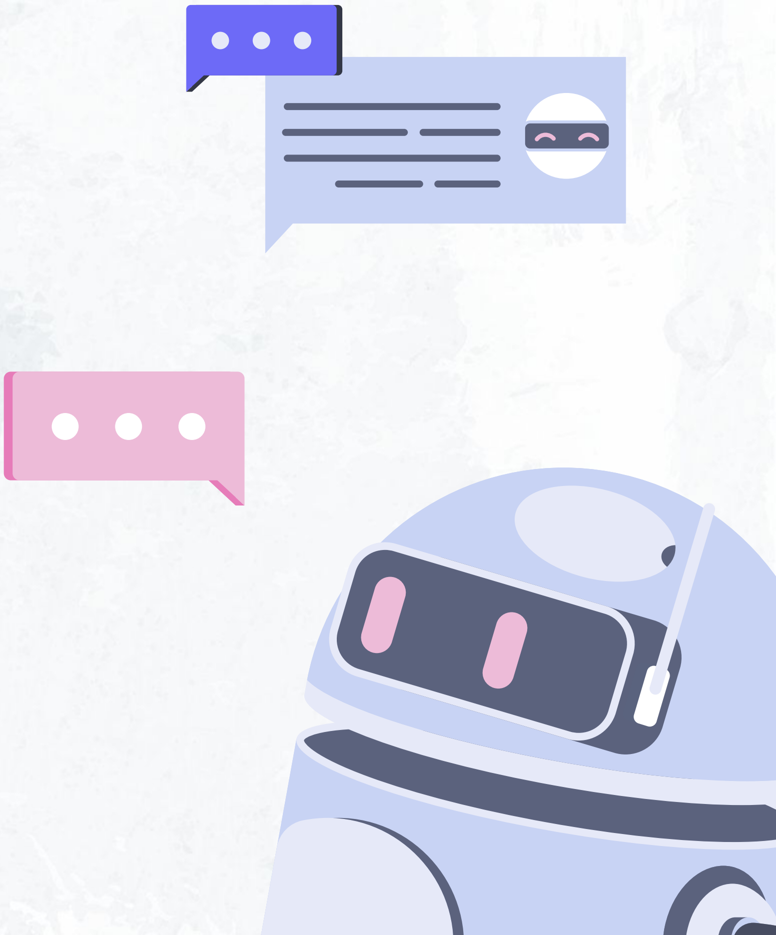
Setrem

Introdução à Inteligência Artificial

João Paulo Aires

(IA)



Índice

01 →

Regressão Linear

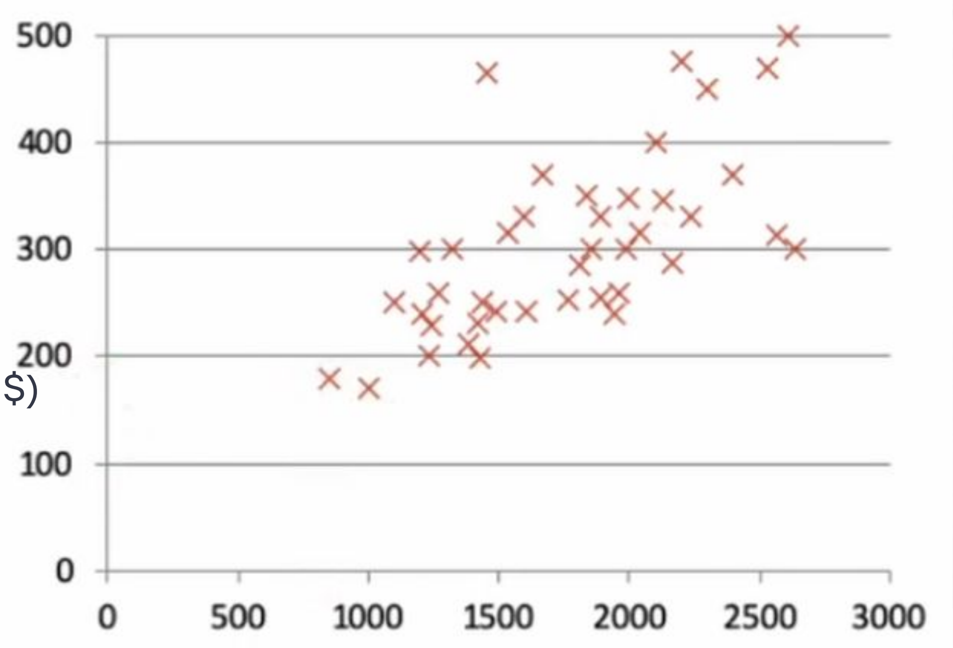
(IA)

Regressão Linear

Preços de Casas (Portland)

Tamanho em feet² (x)	Preço ($) em milhares ($f(x)$)
2104	460
1416	232
1534	315
852	178
...	...

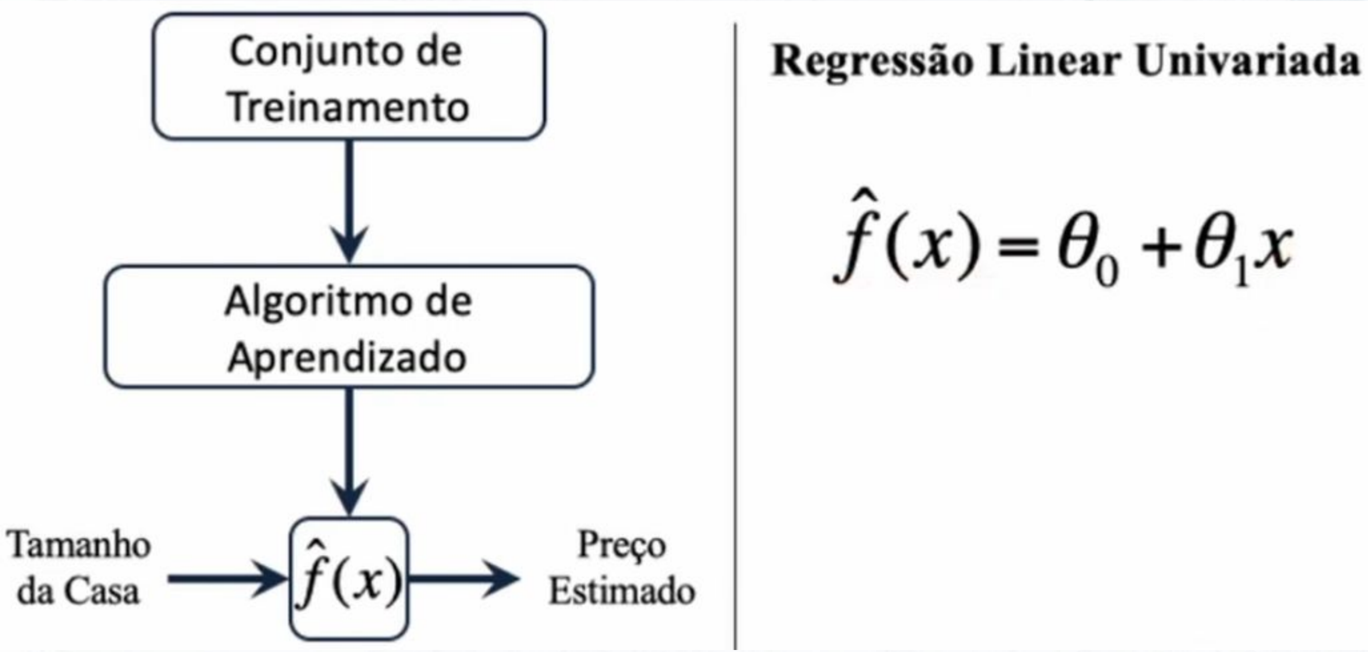
Preço
(em milhares $)



Tamanho
(feet^2)

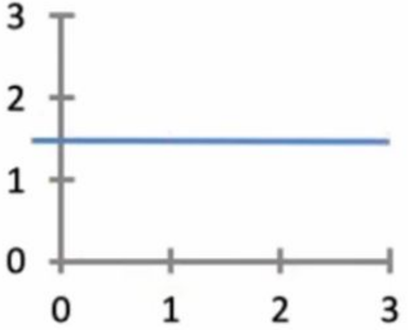
Conjunto de Treinamento

Regressão Linear



Regressão Linear

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

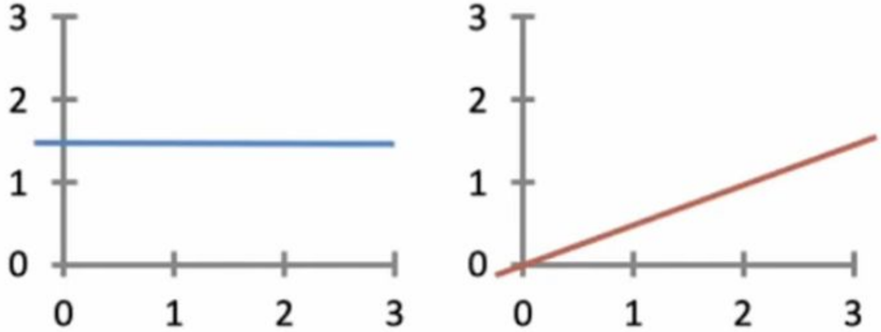


$$\theta_0 = 1.5$$
$$\theta_1 = 0$$

$$\boxed{\hat{f}(x) = 1.5 + 0x}$$

Regressão Linear

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

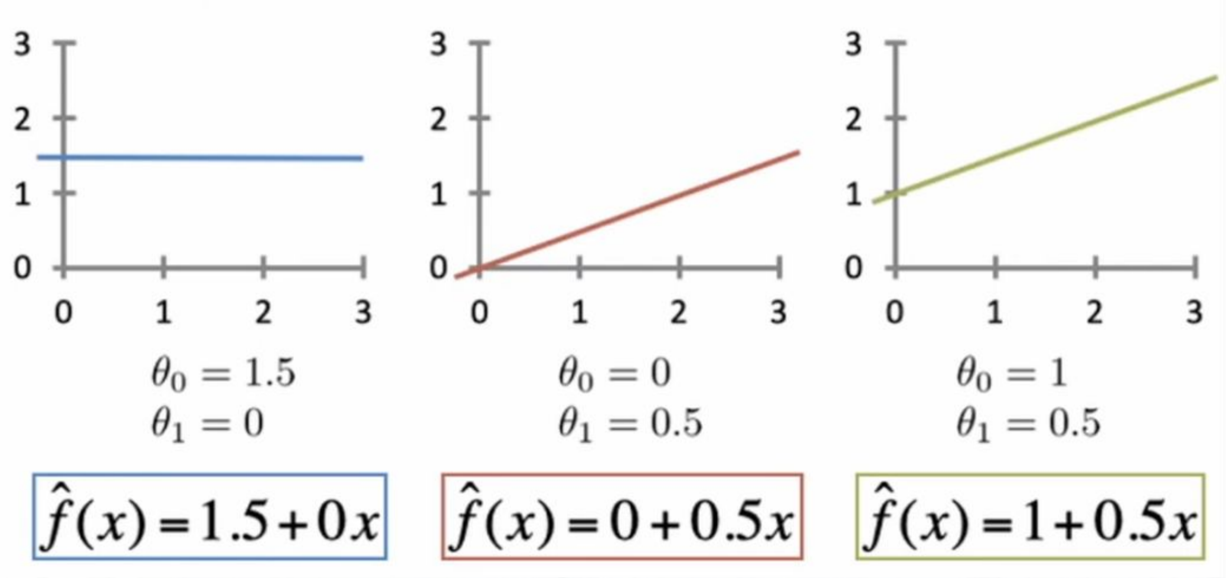


$\theta_0 = 1.5$
$\theta_1 = 0$

$\theta_0 = 0$
$\theta_1 = 0.5$

$$\boxed{\hat{f}(x) = 1.5 + 0x} \qquad \boxed{\hat{f}(x) = 0 + 0.5x}$$

Regressão Linear



Algoritmo Gradiente Descendente

Como minimizar *$J(\theta_0, \theta_1)$*?

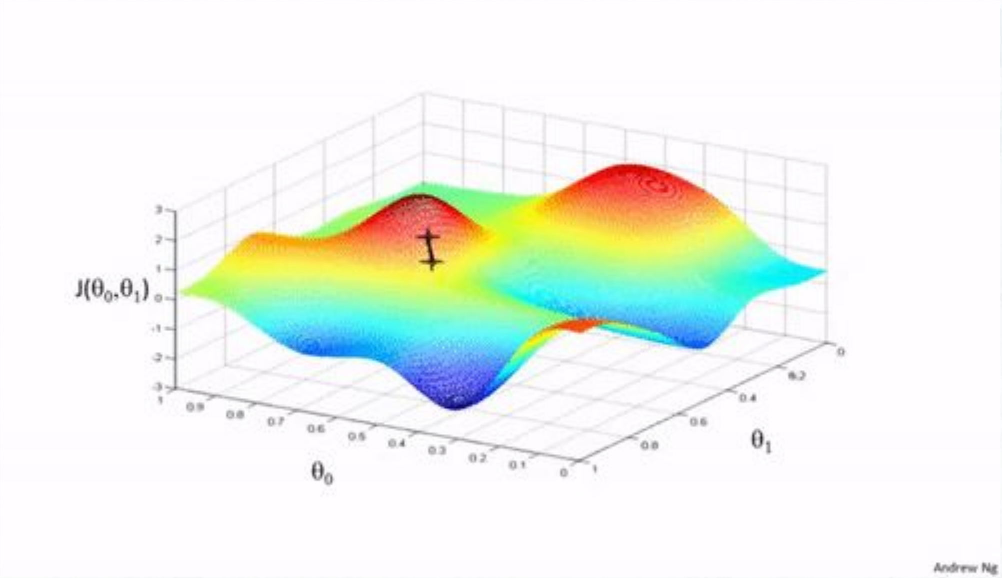
Dada uma função de custo $J(\theta_0, \theta_1)$

Objetivo $\min_{\theta 0,\theta 1} J(\theta_0, \theta_1)$

Algoritmo do Gradiente Descendente:

- Inicialize com algum valor para θ_0, θ_1
- Modifique incrementalmente θ_0, θ_1 para reduzir $J(\theta_0, \theta_1)$ até que possivelmente $J(\theta_0, \theta_1)$ tenha sido minimizada.

Gradiente Descendente



Gradiente Descendente

$$repetir\ até\ convergir\ \{$$

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\}$$

Gradiente Descendente

repetir até convergir {

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

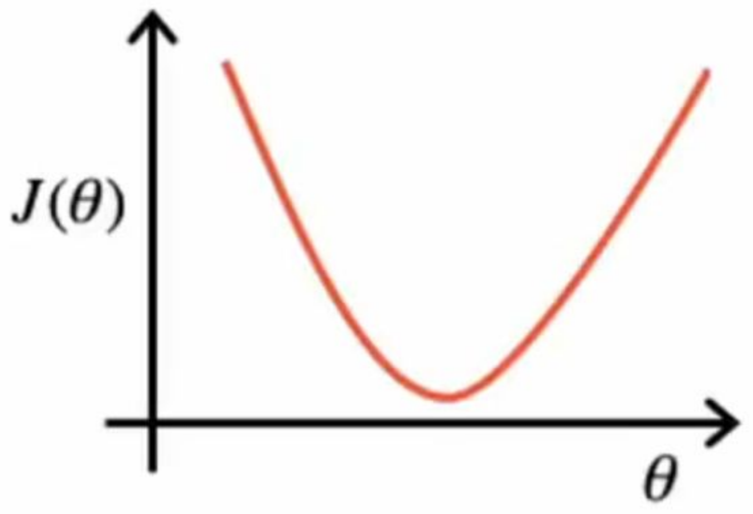
$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

}

Atenção!
Atualização Simultânea!!

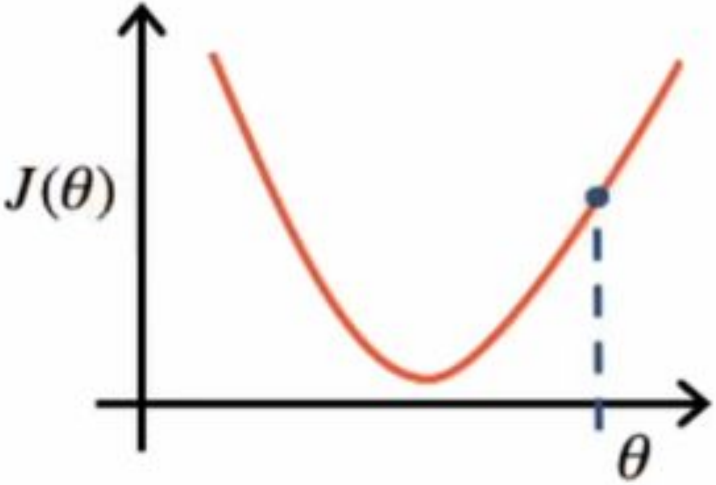
$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Entendendo o papel da derivada



Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

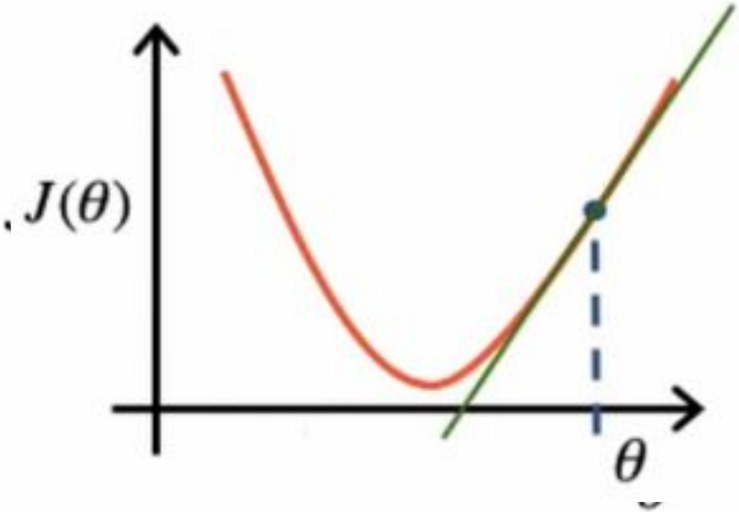
Entendendo o papel da derivada



Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

Entendendo o papel da derivada

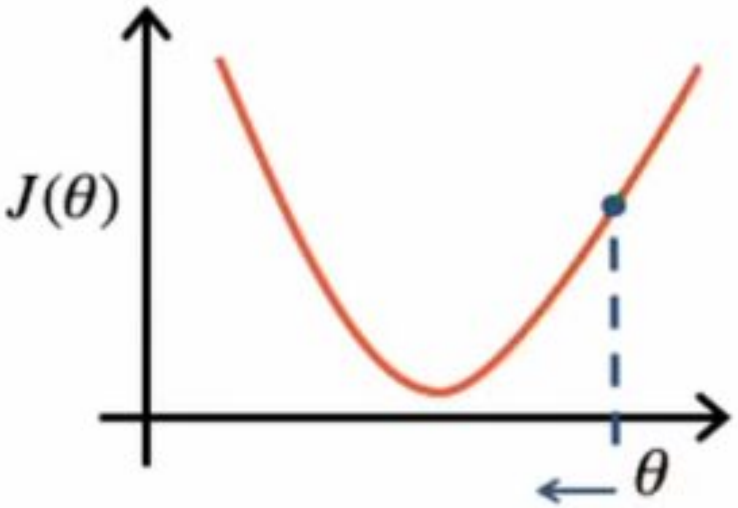


Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

$$\theta := \theta - \alpha \text{ (num positivo)}$$

Entendendo o papel da derivada

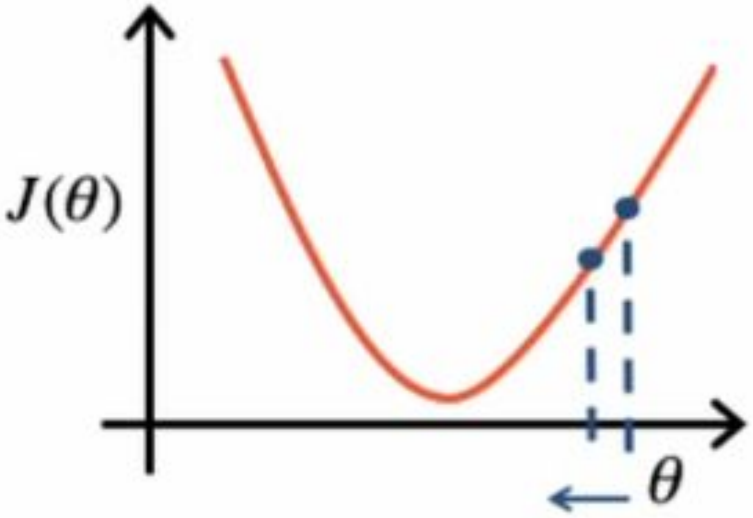


Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

$$\theta := \theta - \alpha \text{ (num positivo)}$$

Entendendo o papel da derivada

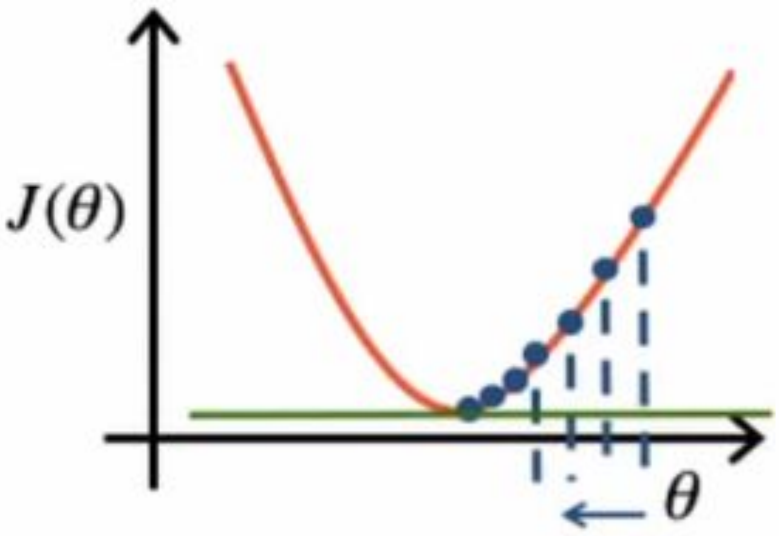


Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

$$\theta := \theta - \alpha \text{ (num positivo)}$$

Entendendo o papel da derivada

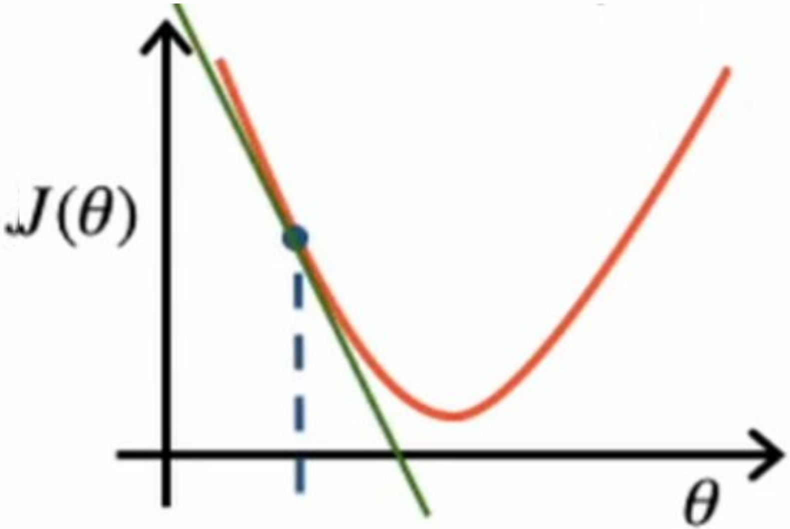


Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

$$\theta := \theta - \alpha \text{ (num positivo)}$$

Entendendo o papel da derivada



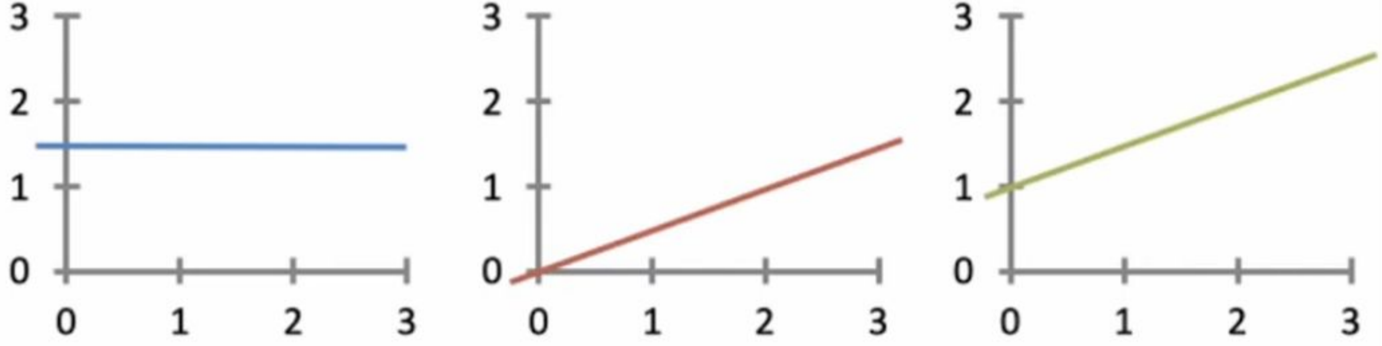
Por simplicidade, assuma a existência de uma função de custo com apenas um parâmetro, $\theta \in \Re$, $J(\theta)$

$$\theta := \theta - \alpha \frac{d}{d\theta} J(\theta)$$

$$\theta := \theta - \alpha \text{ (num negativo)}$$

Regressão Linear

$$\hat{f}(x) = \theta_0 + \theta_1 x$$



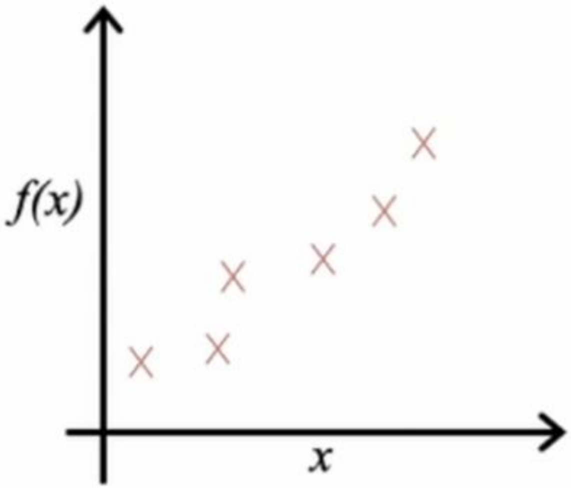
$\theta_0 = 1.5$
$\theta_1 = 0$

$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

$\boxed{\hat{f}(x) = 1.5 + 0x}$ $\boxed{\hat{f}(x) = 0 + 0.5x}$ $\boxed{\hat{f}(x) = 1 + 0.5x}$

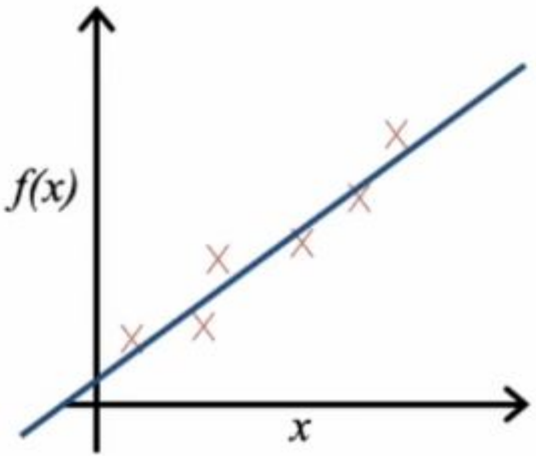
Regressão Linear



$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Ideia: Escolher Θ_0 e Θ_1 de forma que *f'(x)* esteja próximo de *f(x)* para as instâncias de treinamento (*x*, *f(x)*)

Regressão Linear



$$\hat{f}(x) = \theta_0 + \theta_1 x$$

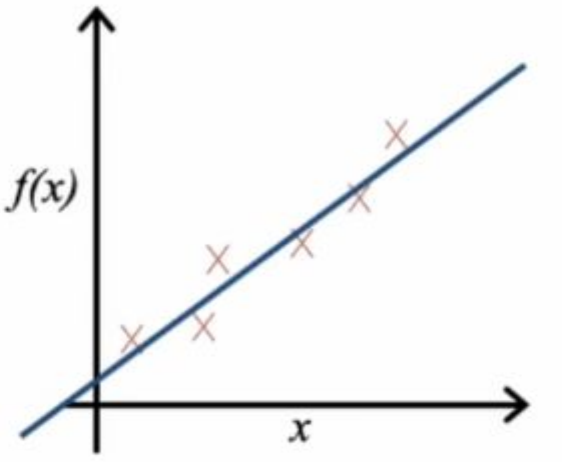
$$\hat{f}(x^{(i)}) - f(x^{(i)})$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)^2$$

$$\frac{1}{N}\sum_{i=1}^{N}\left((\theta_o + \theta_1 x^{(i)}) - f(x^{(i)})\right)^2$$

Ideia: Escolher Θ_o e Θ_1 de forma que *f'(x)* esteja próximo de *f(x)* para as instâncias de treinamento (*x, f(x)*)

Regressão Linear



$$\hat{f}(x) = \theta_0 + \theta_1 x$$

$$\hat{f}(x^{(i)}) - f(x^{(i)})$$

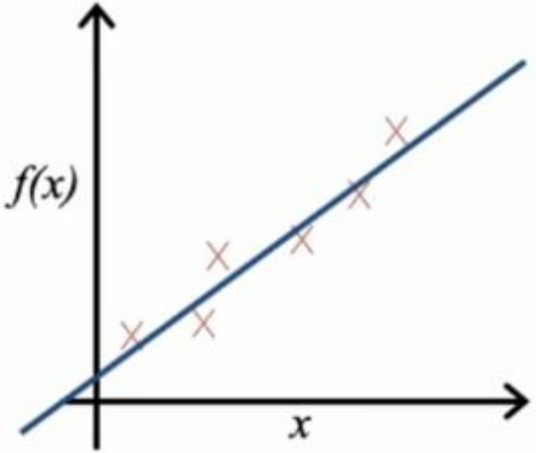
$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)^2$$

$$\frac{1}{N}\sum_{i=1}^{N}\left((\theta_o + \theta_1 x^{(i)}) - f(x^{(i)})\right)^2$$

$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left((\theta_o + \theta_1 x^{(i)}) - f(x^{(i)})\right)^2$$

Ideia: Escolher Θ_o e Θ_1 de forma que *f'(x)* esteja próximo de *f(x)* para as instâncias de treinamento (*x*, *f(x)*)

Regressão Linear



$$\hat{f}(x^{(i)}) - f(x^{(i)})$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)^2$$

$$\frac{1}{N}\sum_{i=1}^{N}\left((\theta_o + \theta_1 x^{(i)}) - f(x^{(i)})\right)^2$$

$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left((\theta_o + \theta_1 x^{(i)}) - f(x^{(i)})\right)^2$$

$$\min_{\theta_o, \theta_1} J(\theta_0, \theta_1)$$

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Ideia: Escolher Θ_o e Θ_1 de forma que *f'(x)* esteja próximo de *f(x)* para as instâncias de treinamento (*x, f(x)*)

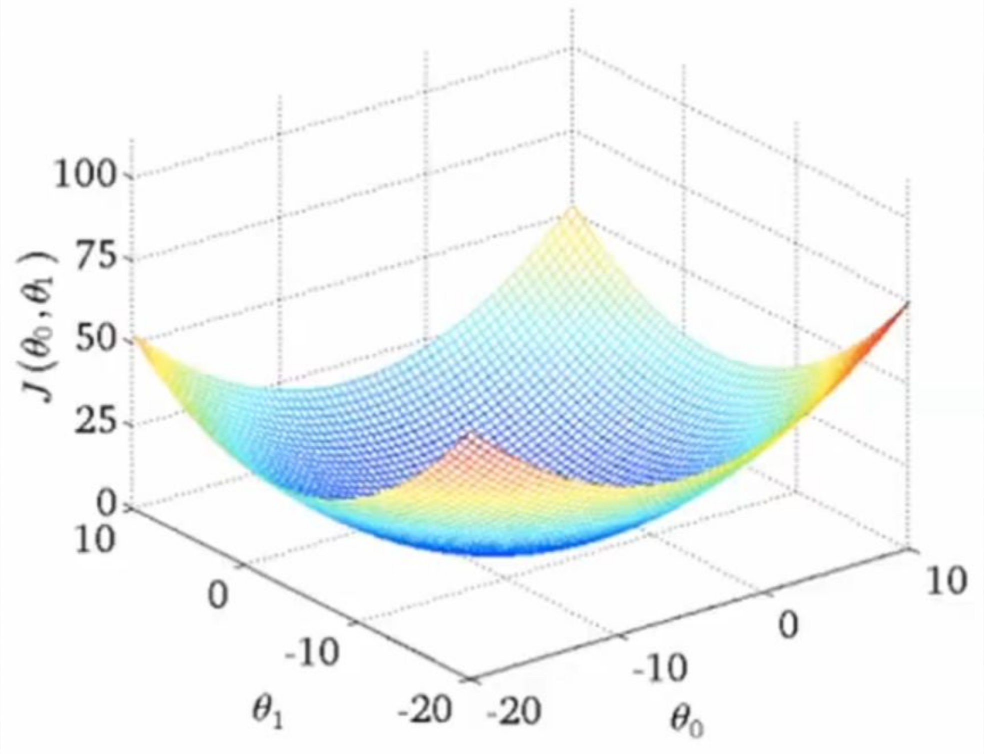
Modelo: $\hat{f}(x) = \theta_0 + \theta_1 x$

Parâmetros: θ_0, θ_1

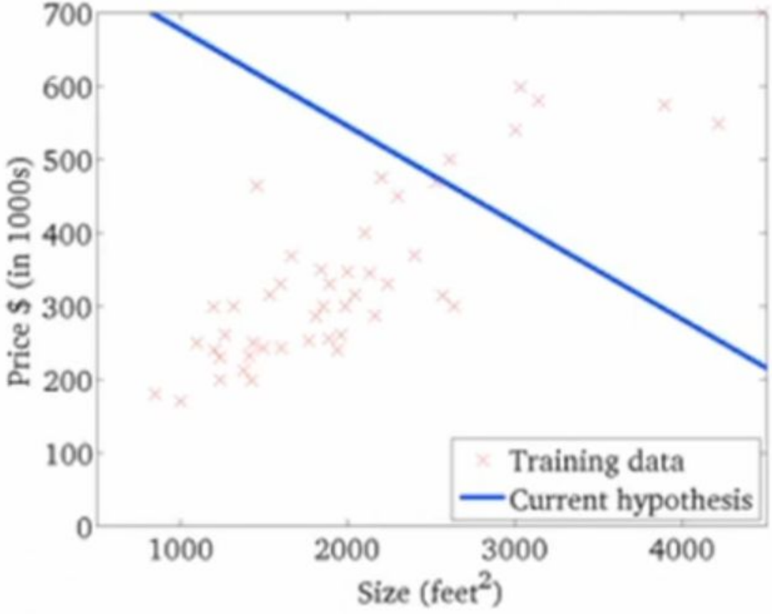
Função de Custo: $J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\hat{f}(x^{(i)}) - f(x^{(i)}) \right)^2$

Objetivo: $\min_{\theta_o, \theta_1} J(\theta_0, \theta_1)$

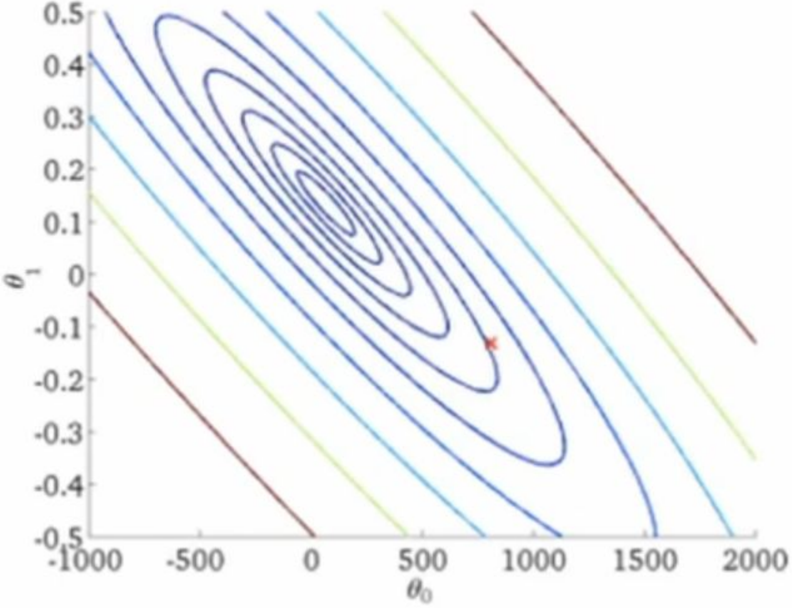
Função de Custo J



f'(x)
(Para valores fixos de θ_0 e θ_1,
é uma função de *x*)

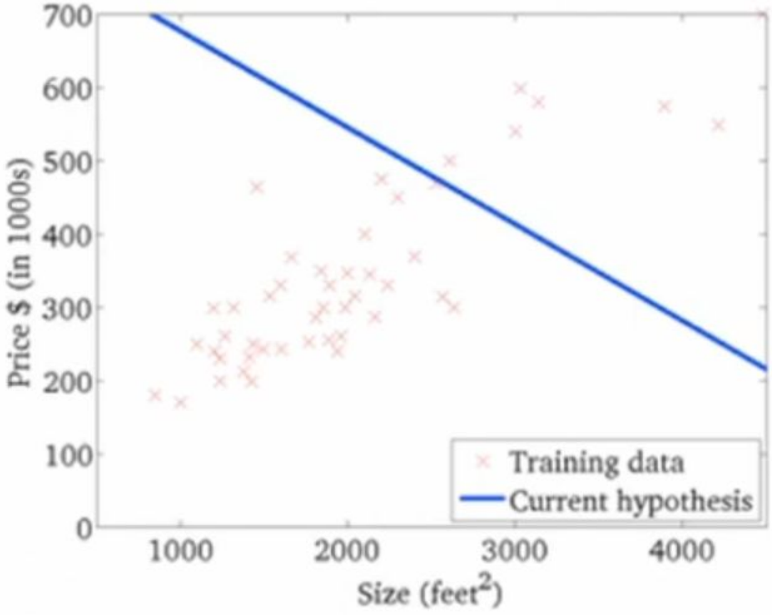


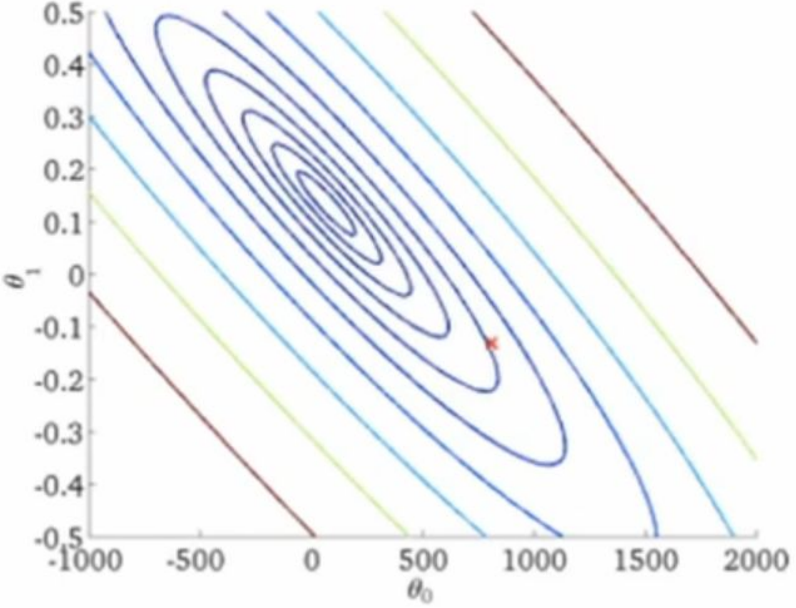
J(θ_0, θ_0)
(Função dos parâmetros θ_0, θ_1)



$f'(x)$
(Para valores fixos de θ_0 e θ_1,
é uma função de *x*)

$J(\theta_0, \theta_0)$
(Função dos parâmetros θ_0, θ_1)

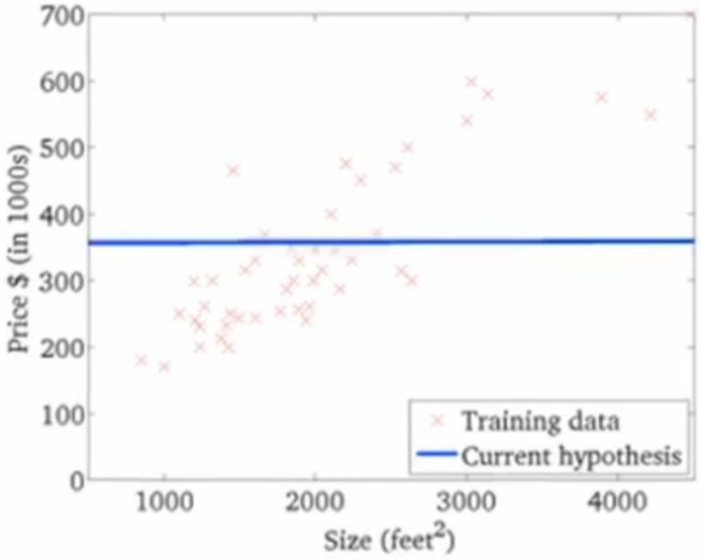


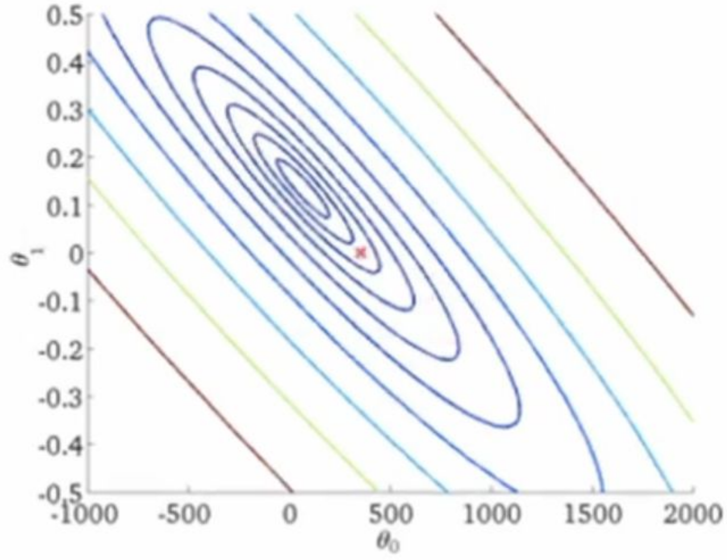


$\theta_0 = 800$
$\theta_1 = -0.15$

f'(x)
(Para valores fixos de θ_0 e θ_1,
é uma função de *x*)

$J(\theta_0, \theta_1)$
(Função dos parâmetros θ_0, θ_1)

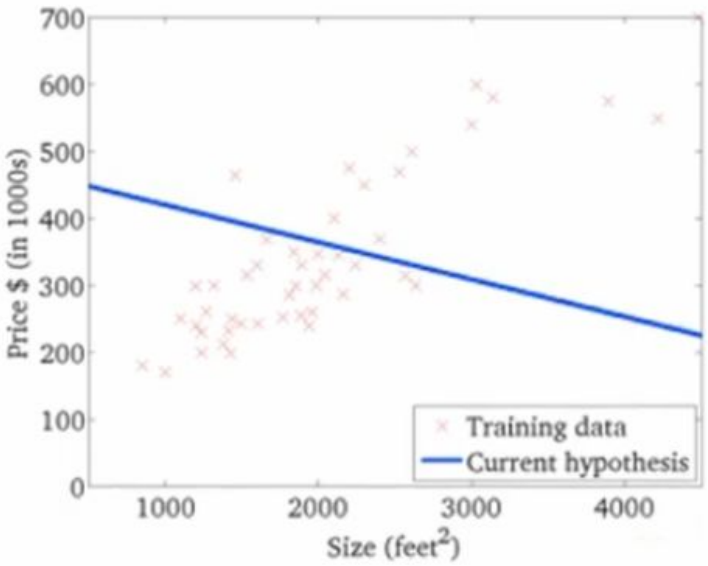


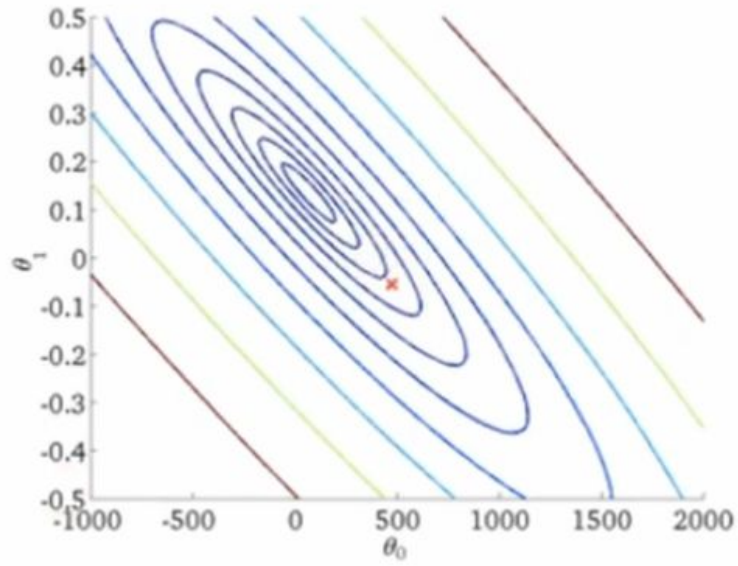


$\theta_0 = 360$
$\theta_1 = 0$

f'(x)
(Para valores fixos de θ_0 e θ_1,
é uma função de *x*)

J(θ_0, θ_1)
(Função dos parâmetros θ_0, θ_1)

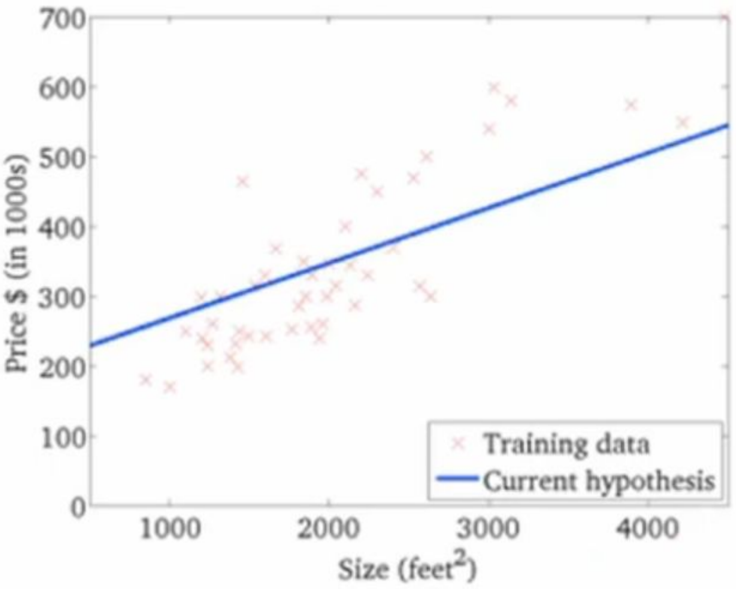


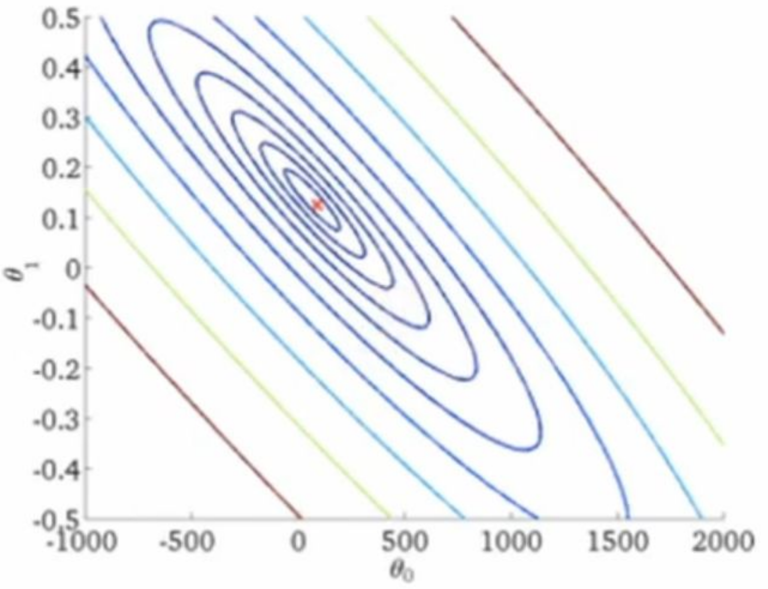


$\theta_0 = 500$
$\theta_1 = -0.025$

$f'(x)$
(Para valores fixos de θ_0 e θ_1,
é uma função de x)

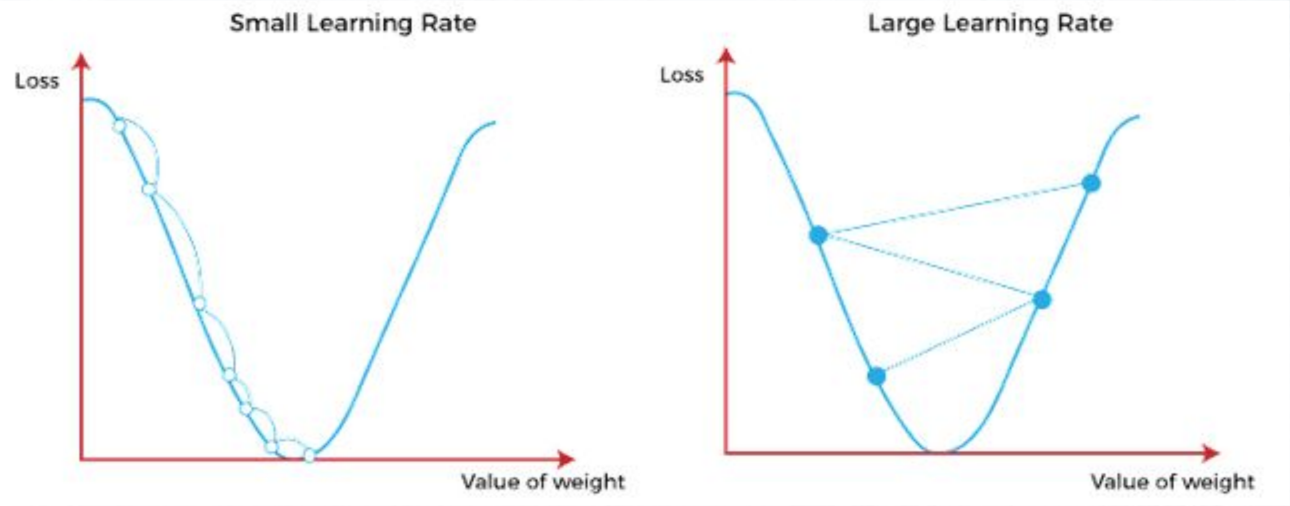
$J(\theta_0, \theta_1)$
(Função dos parâmetros θ_0, θ_1)





$\theta_0 = 230$
$\theta_1 = 0.13$

Tamanho do passo α (taxa de aprendizado)



Calculando as derivadas parciais

$$J(\theta_0,\theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)^2$$

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{1}{N}\sum_{i=1}^{N}\hat{f}(x^{(i)}) - f(x^{(i)}) = \frac{1}{N}\sum_{i=1}^{N}\left(\theta_0 + \theta_1 x^{(i)}\right) - f(x^{(i)})$$

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)x^{(i)} = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - f(x^{(i)})\right)x^{(i)}$$

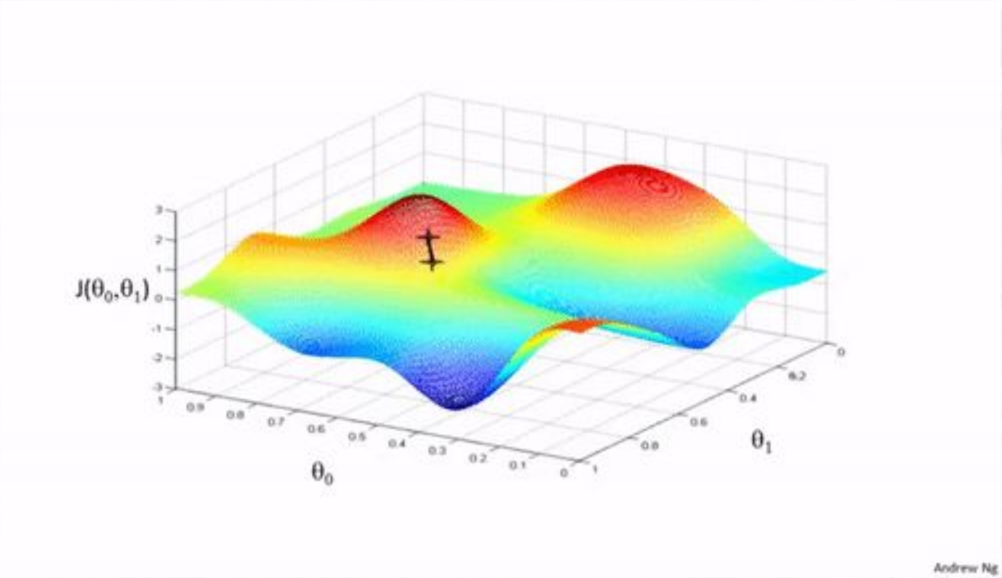
Gradiente Descendente

repetir até convergir {

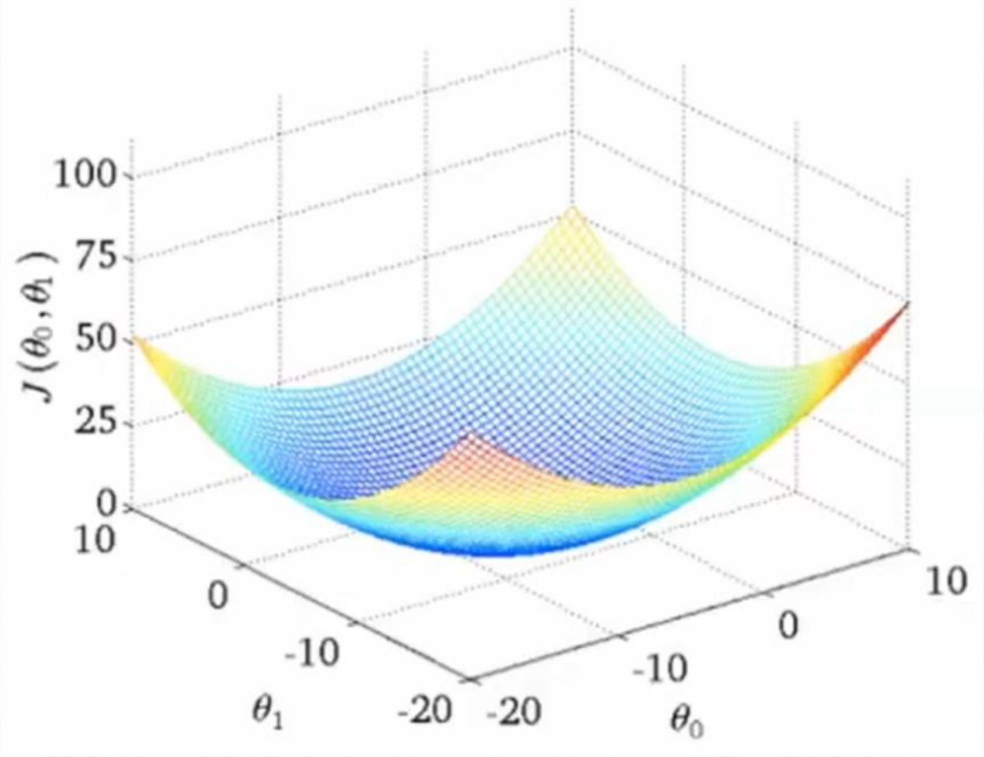
$$\left.\begin{aligned}\theta_0 &:= \theta_0 - \alpha \frac{1}{N}\sum_{i=1}^{N}\left[\left(\theta_0 + \theta_1 x^{(i)}\right) - f(x^{(i)})\right] \\ \theta_1 &:= \theta_1 - \alpha \frac{1}{N}\sum_{i=1}^{N}\left[\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - f(x^{(i)})\right)x^{(i)}\right]\end{aligned}\right\}\text{Atualização Simultânea}$$

}

Gradiente Descendente



Função de Custo: Erro Quadrático Médio



Regressão Linear Univariada

Preços de Casas (Portland)

Tamanho em feet^2 (*x*)	Preço ($) em milhares (*f*(*x*))
2104	460
1416	232
1534	315
852	178
…	…

Conjunto de Treinamento

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Regressão Linear Multivariada

Mas e se tivermos...

Tamanho em feet^2 (x)	# Quartos	# Andares	Anos	Preço ($) em milhares ($f(x)$)
2104	5	1	45	460
1416	3	2	40	232
1534	3	4	30	315
852	2	1	36	178
...	...	...	...	...

Regressão Linear Multivariada

Notação

$m =$ número de atributos

$\mathbf{x}^{(i)} =$ i-ésima instância → ex: $\mathbf{x}^{(3)} = \begin{bmatrix} 1534 \\ 3 \\ 4 \\ 30 \\ 315 \end{bmatrix}$

$x_j^{(i)} =$ valor do j-ésimo atributo da i-ésima instância → ex: $x_3^{(2)} = 4$

Tam	# Q	# A	Anos	Preç
2104	5	1	45	460
1416	3	2	40	232
1534	3	4	30	315
852	2	1	36	178
…	…	…	…	…

Regressão Univariada

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Regressão Multi-variada

$$\hat{f}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$$

$$\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$$

Notação Vetorial para Regressão Multi-Variada

$$\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$$

$$x_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad N \times 1$$

Notação Vetorial para Regressão Multi-Variada

$$\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$$

$$x_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad N \times 1$$

Logo, $\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (m+1) \times 1$

Notação Vetorial para Regressão Multi-Variada

$$\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$$

$$x_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad N \times 1$$

Logo, $\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$ $(m+1) \times 1$

$$\underset{(m+1)\times 1}{\Theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$$

Notação Vetorial para Regressão Multi-Variada

Podemos reescrever $\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$

como $\hat{f}(\mathbf{x}) = \Theta^T \mathbf{x}$

Notação Vetorial para Regressão Multi-Variada

Podemos reescrever $\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$

como $\hat{f}(\mathbf{x}) = \Theta^T \mathbf{x}$

$$\Theta^T = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_m \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Gradiente Descendente para Regressão Multi-Variada

Modelo: $\hat{f}(\mathbf{x}) = \Theta^T \mathbf{x}$

Parâmetros: Θ

Função de Custo: $J(\Theta) = \dfrac{1}{2N} \sum_{i=1}^{N} \left(\Theta^T \mathbf{x}^{(i)} - f(\mathbf{x}^{(i)})\right)^2$

Gradiente Descendente para Regressão Multi-Variada

Anteriormente ($m = 1$)

repetir até convergir {

$$\theta_0 := \theta_0 - \alpha \frac{1}{N} \sum_{i=1}^{N} \left[\left(\theta_0 + \theta_1 x^{(i)} \right) - f(x^{(i)}) \right]$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{N} \sum_{i=1}^{N} \left[\left(\left(\theta_0 + \theta_1 x^{(i)} \right) - f(x^{(i)}) \right) x^{(i)} \right]$$

}

Novo Algoritmo ($m \geq 1$)

repetir até convergir {

$$\theta_j := \theta_j - \alpha \frac{1}{N} \sum_{i=1}^{N} \left[\left(\Theta^T \mathbf{x}^{(i)} - f(\mathbf{x}^{(i)}) \right) x_j^{(i)} \right]$$

para $(j = 0, 1, \ldots, m)$

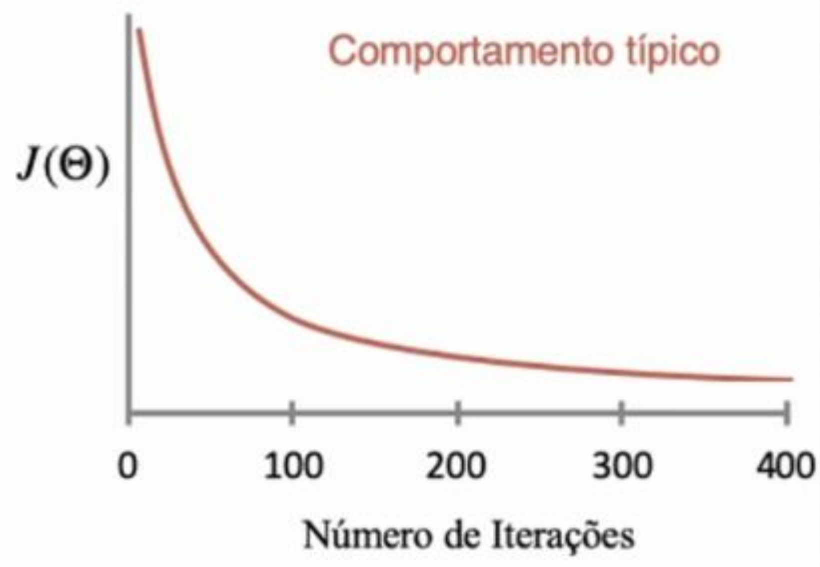
}

Gradiente Descendente para Regressão Multi-Variada

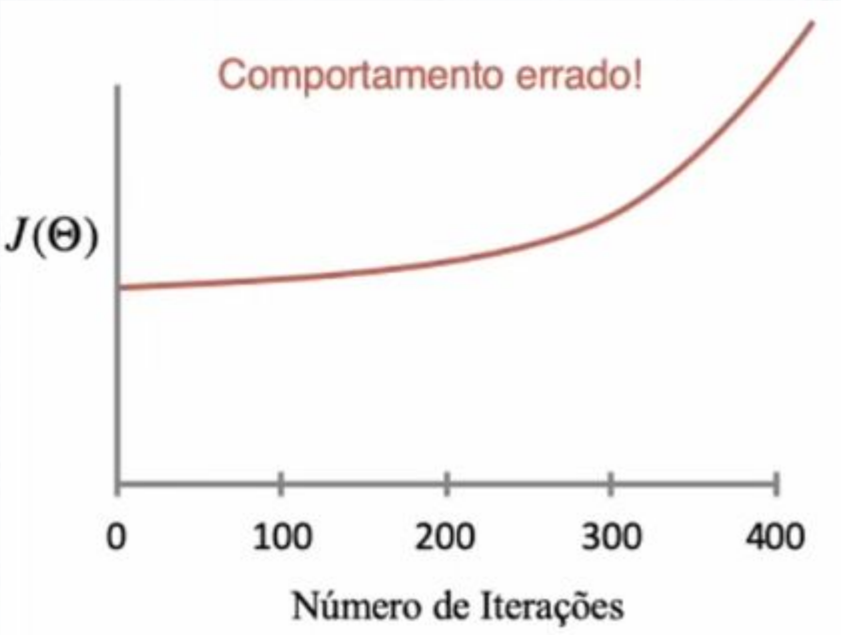
Dicas:

- Normalizar os atributos para acelerar convergência
  - Utilizar estratégias já vistas como padronização ou transformação para intervalos [0, 1] ou [-1, 1]
  - Garantir que o gradiente descendente esteja funcionando
    - Plotar função de custo x iteração do gradiente descendente

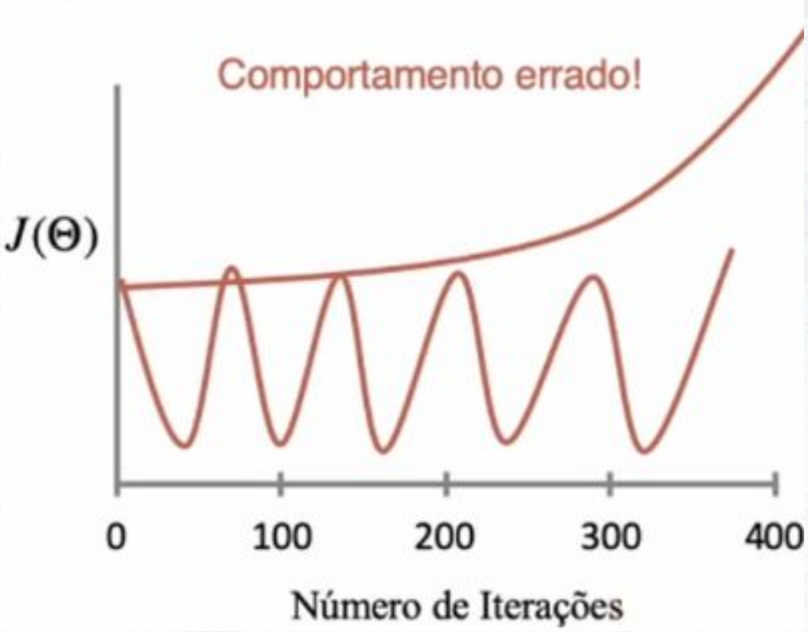
Gradiente Descendente para Regressão Multi-Variada



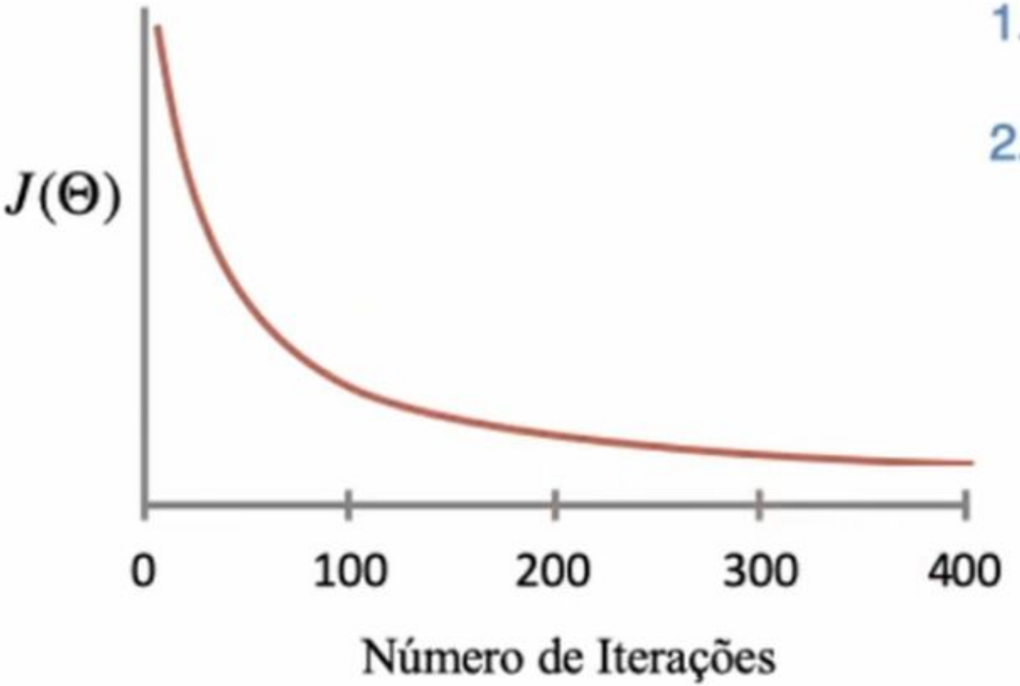
Gradiente Descendente para Regressão Multi-Variada



Gradiente Descendente para Regressão Multi-Variada



Debugando o Gradiente Descendente



Heurística para escolha de α:

1. Começar com valores pequenos (ex: 0.001)
2. Incrementar o valor por algum fator (ex: 3, 10, etc.) para agilizar convergência, mas sempre conferindo se os valores estão decrescendo após cada iteração

02 →

Regressão Logística

(IA)

Regressão Logística

- Apesar do nome, é um algoritmo de classificação!

Regressão Logística

- Apesar do nome, é um algoritmo de classificação!
- Utilizado para discriminar entre duas classes
  - $f(x) = \{0, 1\}$

Regressão Logística

- Apesar do nome, é um algoritmo de classificação!
- Utilizado para discriminar entre duas classes
  - $f(x) = \{0, 1\}$
  - Geralmente, 0 indica ausência (classe negativa) e 1 indica presença (classe positiva) do que se deseja classificar

Regressão Logística

- Apesar do nome, é um algoritmo de classificação!
- Utilizado para discriminar entre duas classes
    - $f(x) = \{0, 1\}$
    - Geralmente, 0 indica ausência (classe negativa) e 1 indica presença (classe positiva) do que se deseja classificar
        - Ex: diagnóstico de HIV
            - 0 = sem HIV (classe negativa)
            - 1 = com HIV (classe positiva)

Regressão Logística

- Gera função $0 <= f'(x) <= 1$
- Em regressão linear, $f'(x) = \Theta^T\mathbf{x}$
- Qual o modelo para regressão logística?

Regressão Logística

- Gera função $0 <= f'(x) <= 1$
- Em regressão linear, $f'(x) = \Theta^T \mathbf{x}$
- Qual o modelo para regressão logística?
  - $f'(x) = sigmoide(\Theta^T \mathbf{x})$

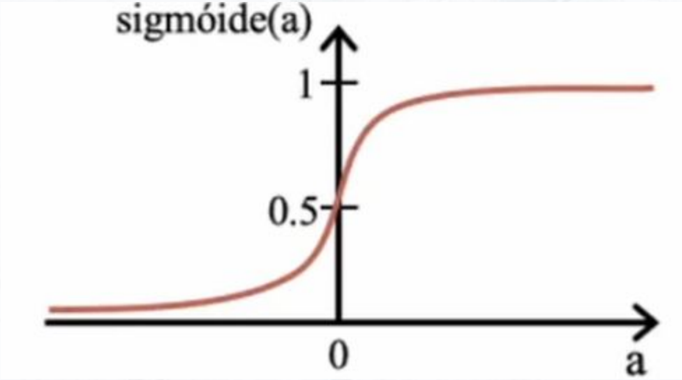
Regressão Logística

- Gera função 0 <= $f'(x)$ <= 1
- Em regressão linear, $f'(x) = \Theta^T \mathbf{x}$
- Qual o modelo para regressão logística?
  - $f'(x) = sigmoide(\Theta^T \mathbf{x})$

$$sigmoide(a) = \frac{1}{1 + e^{-a}}$$

Regressão Logística

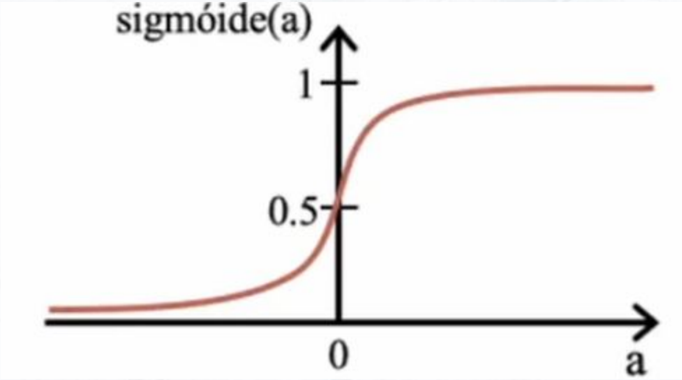
- Gera função 0 <= $f'(x)$ <= 1
- Em regressão linear, $f'(x) = \Theta^T \mathbf{x}$
- Qual o modelo para regressão logística?
    - $f'(x) = \mathit{sigmoide}(\Theta^T \mathbf{x})$



$$\mathit{sigmoide}(a) = \frac{1}{1 + e^{-a}}$$

Regressão Logística

- Gera função $0 <= f'(x) <= 1$
- Em regressão linear, $f'(x) = \Theta^T \mathbf{x}$
- Qual o modelo para regressão logística?
    - $f'(x) = sigmoide(\Theta^T \mathbf{x})$



$$sigmoide(a) = \frac{1}{1+e^{-a}}$$

$$\boxed{\hat{f}(\mathbf{x}) = \frac{1}{1+e^{-(\Theta^T \mathbf{x})}}}$$

Regressão Logística

- Interpretação **probabilística**
  - **Probabilidade estimada da classe positiva**

Regressão Logística

- Interpretação **probabilística**
  - **Probabilidade estimada da classe positiva**
    - Ex: classificação de tumor: $f(x)$ = {benigno, maligno}
    - Para paciente com $f'(\mathbf{x}) = 0.7$
      - Paciente tem 70% de chance de ter tumor maligno

Regressão Logística

- Interpretação **probabilística**
  - **Probabilidade estimada da classe positiva**
    - Ex: classificação de tumor: $f(x)$ = {benigno, maligno}
    - Para paciente com $f'(\mathbf{x}) = 0.7$
      - Paciente tem 70% de chance de ter tumor maligno

$$\hat{f}(\mathbf{x}) = p(f(\mathbf{x}) = 1 \mid \mathbf{x}; \Theta)$$

"Probabilidade que $f(\mathbf{x}) = 1$, dado **x** parametrizado por Θ"

Regressão Logística

- Interpretação **probabilística**
  - **Probabilidade estimada da classe positiva**
    - Ex: classificação de tumor: *f*(*x*) = {benigno, maligno}
    - Para paciente com *f'*(**x**) = 0.7
      - Paciente tem 70% de chance de ter tumor maligno

$$\hat{f}(\mathbf{x}) = p(f(\mathbf{x}) = 1 \mid \mathbf{x}; \Theta)$$

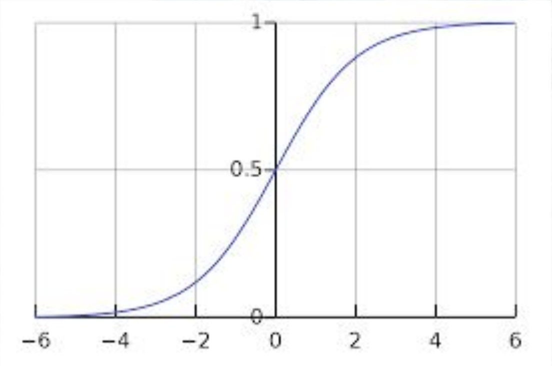
"Probabilidade que *f*(**x**) = 1, dado **x** parametrizado por Θ"

$$p(f(\mathbf{x}) = 0 \mid \mathbf{x}; \Theta) = 1 - p(f(\mathbf{x}) = 1 \mid \mathbf{x}; \Theta)$$

Regressão Logística

- Dada a interpretação probabilística da função logística, o que realmente estamos fazendo é:
  - *se sigmoide*($\Theta^T \mathbf{x}$) >= 0.5
    - então $\mathbf{x}$ é da classe positiva
    - senão $\mathbf{x}$ é da classe negativa

sigmoide($\Theta^T \mathbf{x}$)

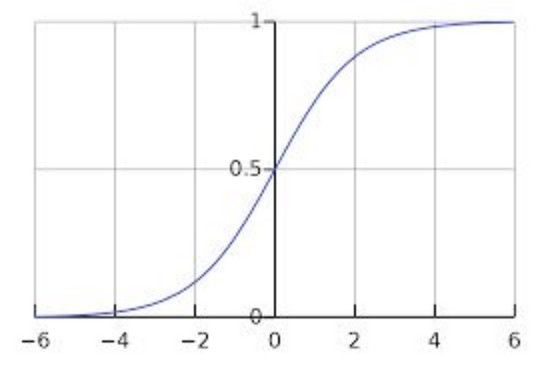


$\Theta^T \mathbf{x}$

Regressão Logística

- Dada a interpretação probabilística da função logística, o que realmente estamos fazendo é:
  - *se sigmoide*(Θ^T**x**) >= 0.5
    - então **x** é da classe positiva
    - senão **x** é da classe negativa
- O que isso nos diz a respeito de Θ^T**x**?
  - se Θ^T**x** >= 0
    - então **x** é da classe positiva
    - senão **x** é da classe negativa

sigmoide(Θ^T**x**)

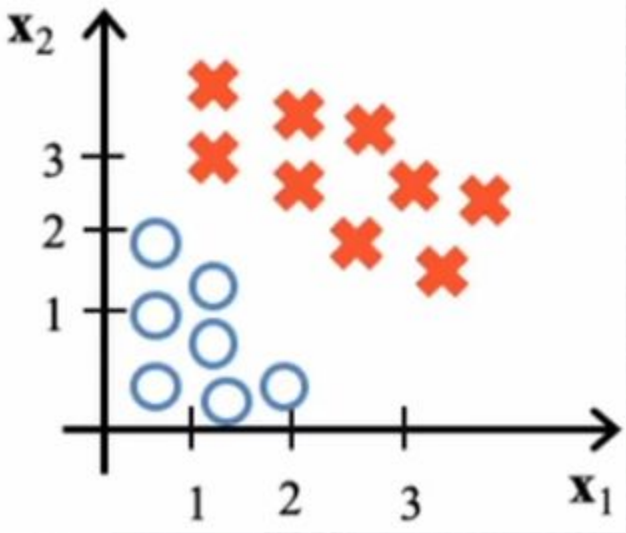


Θ^T**x**

Regressão Logística

- Ex: Problema com dois atributos

$$\hat{f}(\mathbf{x}) = sigmoide(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

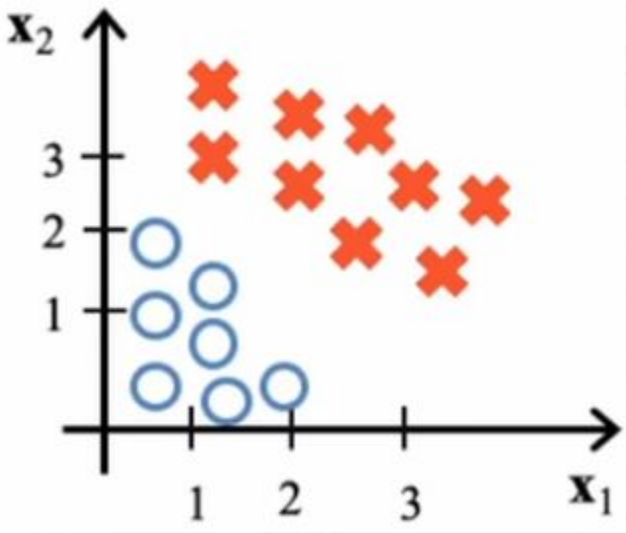


Regressão Logística

- Ex: Problema com dois atributos

$$\hat{f}(\mathbf{x}) = sigmoide(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

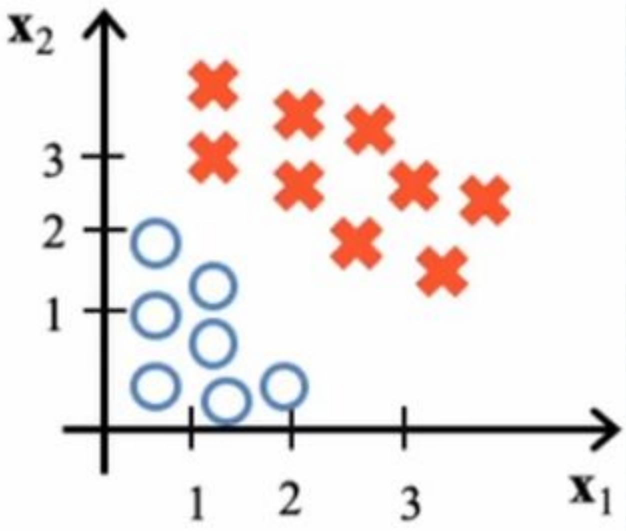
$$\Theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$



Regressão Logística

- Ex: Problema com dois atributos

$$\hat{f}(\mathbf{x}) = sigmoide(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$



$$\Theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

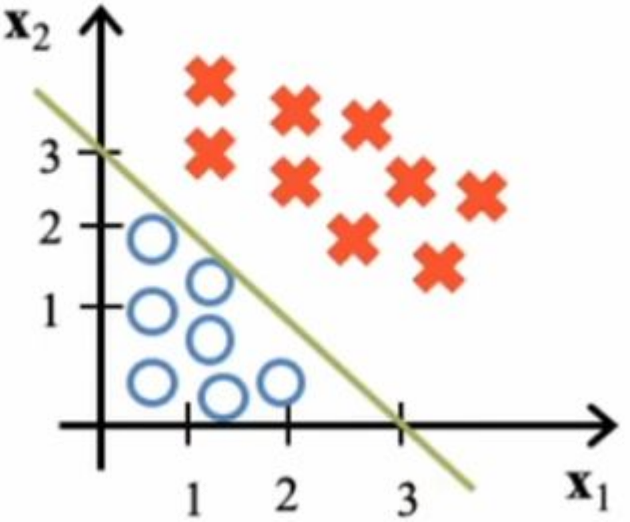
$$classe\ positiva: \; -3 + 1x_1 + 1x_2 \geq 0$$

$$x_1 + x_2 \geq 3$$

Regressão Logística

- Ex: Problema com dois atributos

$$\hat{f}(\mathbf{x}) = sigmoide(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$



$$\Theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$classe\ positiva: -3 + 1x_1 + 1x_2 \geq 0$$

$$x_1 + x_2 \geq 3$$

$x_1 + x_2 = 3$

Regressão Logística

- O problema de otimização novamente se resume a minimizar uma função de custo
- Função de custo quadrática da regressão linear pode ser usada para regressão logística?

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\hat{f}(x^{(i)}) - f(x^{(i)}) \right)^2$$

Regressão Logística

- O problema de otimização novamente se resume a minimizar uma função de custo
- Função de custo quadrática da regressão linear pode ser usada para regressão logística?

$$J(\theta_0,\theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\hat{f}(x^{(i)}) - f(x^{(i)})\right)^2$$

Não! Pois $J(\Theta)$ vira uma função não-convexa!!
Logo possui vários ótimos locais!

Regressão Logística

- Função de custo a ser minimizada:

$$J(\Theta) = \frac{1}{N}\sum_{i=1}^{N} custo(\hat{f}(\mathbf{x}^{(i)}), f(\mathbf{x}^{(i)}))$$

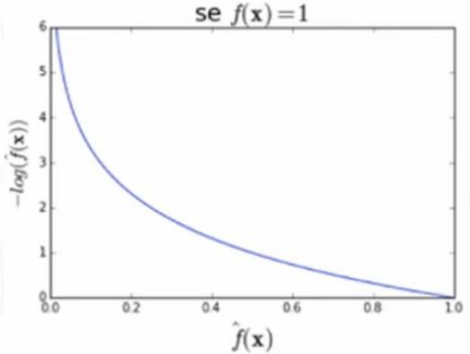
$$custo(\hat{f}(\mathbf{x}), f(\mathbf{x})) = \begin{cases} -\log(\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 1 \\ -\log(1-\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 0 \end{cases}$$

Regressão Logística

- Função de custo a ser minimizada:

$$J(\Theta) = \frac{1}{N} \sum_{i=1}^{N} custo(\hat{f}(\mathbf{x}^{(i)}), f(\mathbf{x}^{(i)}))$$

$$custo(\hat{f}(\mathbf{x}), f(\mathbf{x})) = \begin{cases} -\log(\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 1 \\ -\log(1 - \hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 0 \end{cases}$$

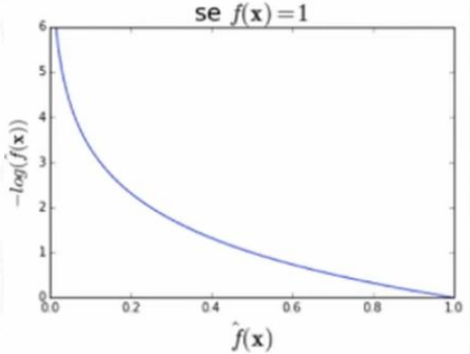


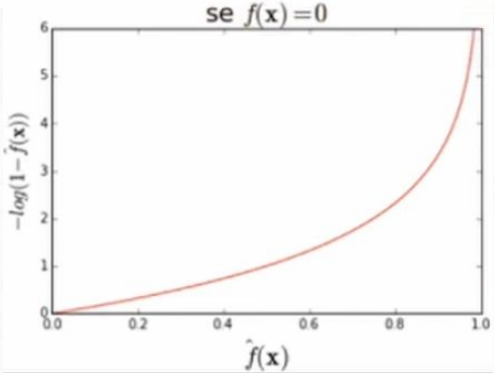
Regressão Logística

- Função de custo a ser minimizada:

$$J(\Theta) = \frac{1}{N}\sum_{i=1}^{N} custo(\hat{f}(\mathbf{x}^{(i)}), f(\mathbf{x}^{(i)}))$$

$$custo(\hat{f}(\mathbf{x}), f(\mathbf{x})) = \begin{cases} -\log(\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 1 \\ -\log(1-\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 0 \end{cases}$$





Regressão Logística

- Reescrevendo a função de custo:

$$J(\Theta) = \frac{1}{N}\sum_{i=1}^{N} custo(\hat{f}(\mathbf{x}^{(i)}), f(\mathbf{x}^{(i)}))$$

$$custo(\hat{f}(\mathbf{x}), f(\mathbf{x})) = \begin{cases} -\log(\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 1 \\ -\log(1-\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 0 \end{cases}$$

$$J(\Theta) = \frac{1}{N}\sum_{i=1}^{N} -f(\mathbf{x}^{(i)})\log(\hat{f}(\mathbf{x}^{(i)})) - (1 - f(\mathbf{x}^{(i)}))\log(1-\hat{f}(\mathbf{x}^{(i)}))$$

Regressão Logística

- Reescrevendo a função de custo:

$$J(\Theta) = \frac{1}{N} \sum_{i=1}^{N} custo(\hat{f}(\mathbf{x}^{(i)}), f(\mathbf{x}^{(i)}))$$

$$custo(\hat{f}(\mathbf{x}), f(\mathbf{x})) = \begin{cases} -\log(\hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 1 \\ -\log(1 - \hat{f}(\mathbf{x})) & se\ f(\mathbf{x}) = 0 \end{cases}$$

$$J(\Theta) = \frac{1}{N} \sum_{i=1}^{N} -f(\mathbf{x}^{(i)}) \log(\hat{f}(\mathbf{x}^{(i)})) - (1 - f(\mathbf{x}^{(i)})) \log(1 - \hat{f}(\mathbf{x}^{(i)}))$$

$$J(\Theta) = -\frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}^{(i)}) \log(\hat{f}(\mathbf{x}^{(i)})) + (1 - f(\mathbf{x}^{(i)})) \log(1 - \hat{f}(\mathbf{x}^{(i)}))$$

Regressão Logística

- Sabendo que a nova função de custo é **convexa**, como minimizá-la?
  - Gradiente Descendente!

$$J(\Theta) = -\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}^{(i)})\log(\hat{f}(\mathbf{x}^{(i)})) + (1 - f(\mathbf{x}^{(i)}))\log(1 - \hat{f}(\mathbf{x}^{(i)}))$$

Regressão Logística

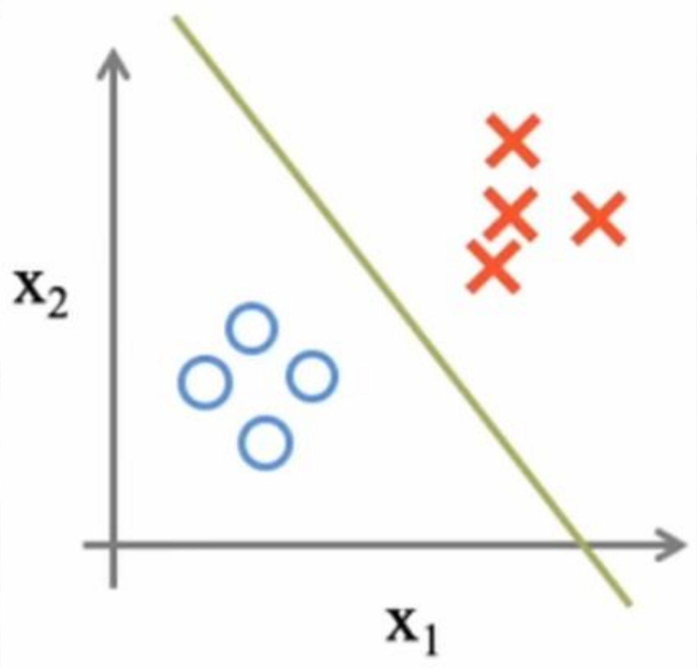
- Sabendo que a nova função de custo é **convexa**, como minimizá-la?
  - Gradiente Descendente!

$$J(\Theta) = -\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}^{(i)})\log(\hat{f}(\mathbf{x}^{(i)})) + (1 - f(\mathbf{x}^{(i)}))\log(1 - \hat{f}(\mathbf{x}^{(i)}))$$

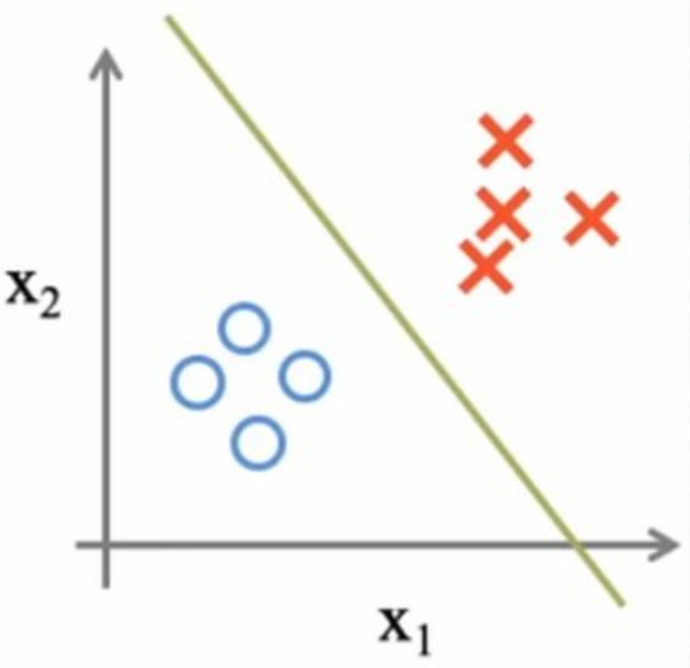
$$\begin{aligned} &\textit{repetir até convergir } \{ \\ &\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\Theta) \\ &\textit{para } (j = 0, 1, \ldots, m) \\ &\} \end{aligned}$$

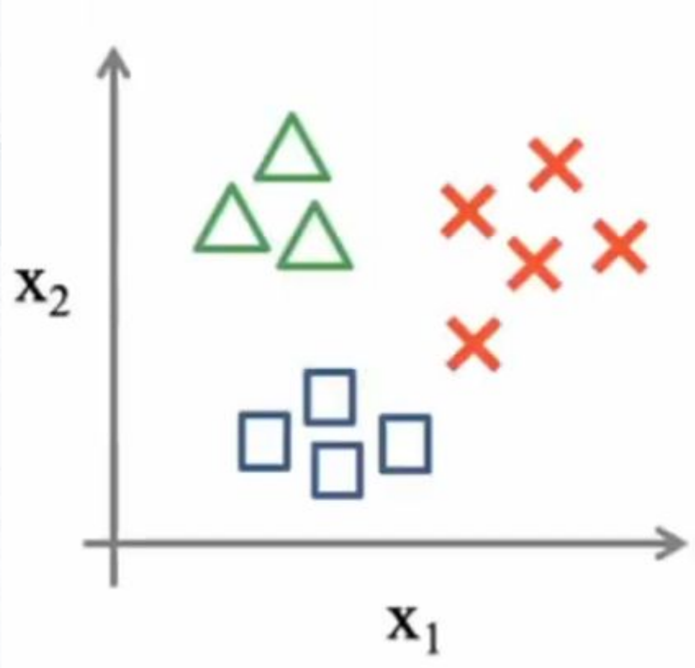
$$\begin{aligned} &\textit{repetir até convergir } \{ \\ &\theta_j := \theta_j - \alpha \frac{1}{N}\sum_{i=1}^{N}\left[\left(\hat{f}(\mathbf{x}^{(i)}) - f(\mathbf{x}^{(i)})\right) x_j^{(i)}\right] \\ &\textit{para } (j = 0, 1, \ldots, m) \\ &\} \end{aligned}$$

Regressão Logística

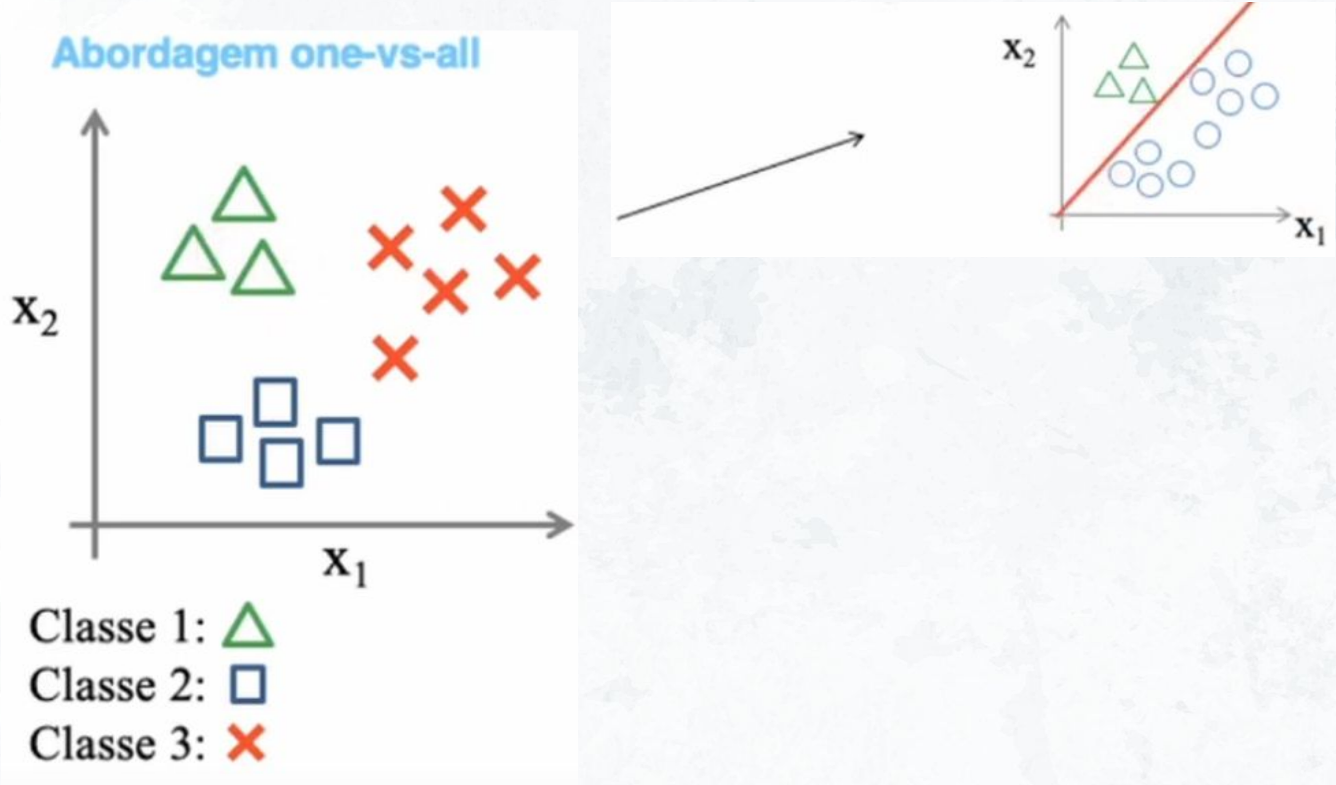


Regressão Logística

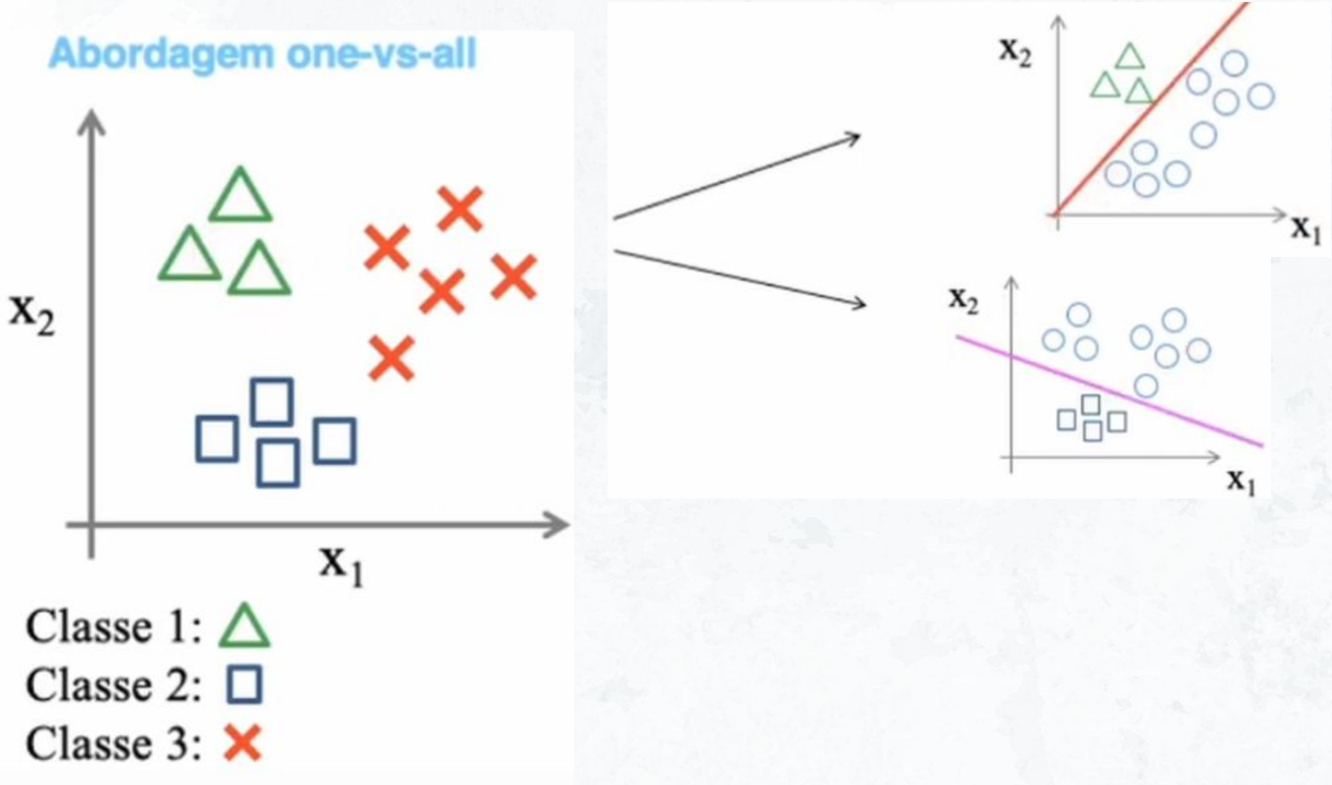




Regressão Logística



Regressão Logística



Regressão Logística