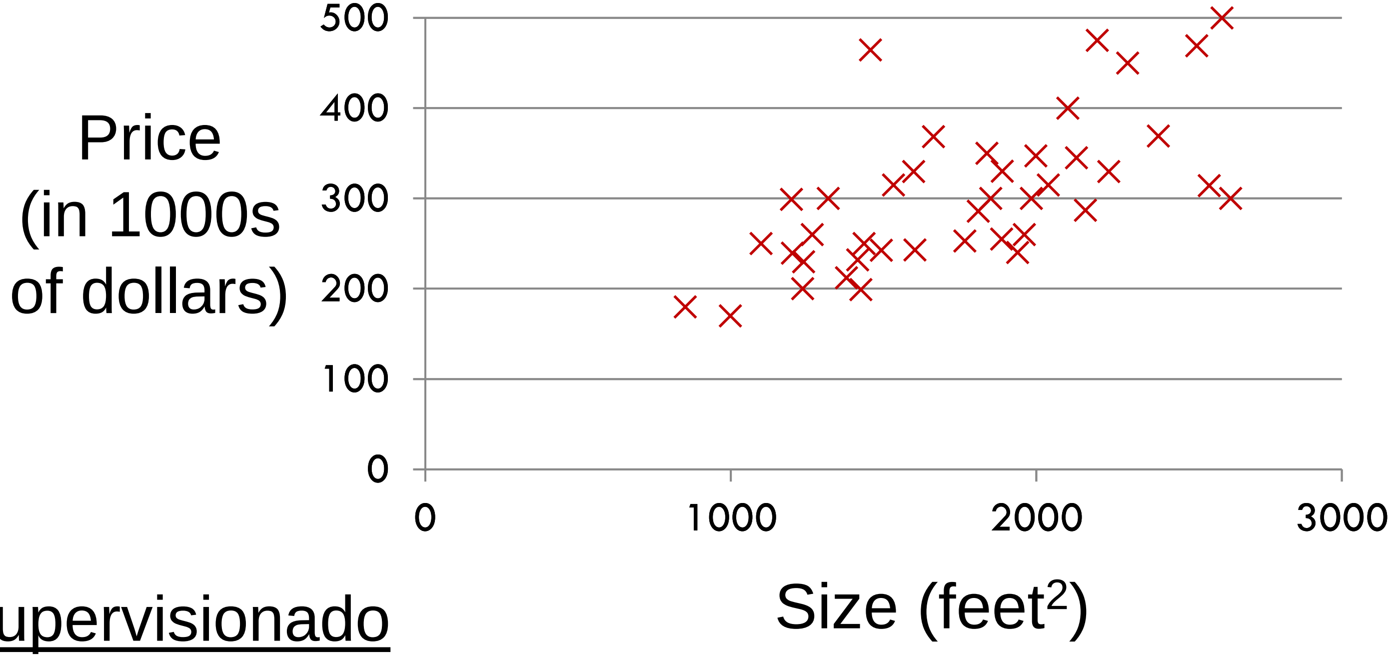
REGRESSÃO LINEAR COM UMA VARIÁVEL

Rogério Gomes e Paulo Almeida

Preços de Casas



Aprendizado Supervisionado

Problema de Regressão
Estima o valor real da saída (contínuo)

Problemas de classificação
Possui valores discretos de saída

Conjunto de treinamento dos preços das casas

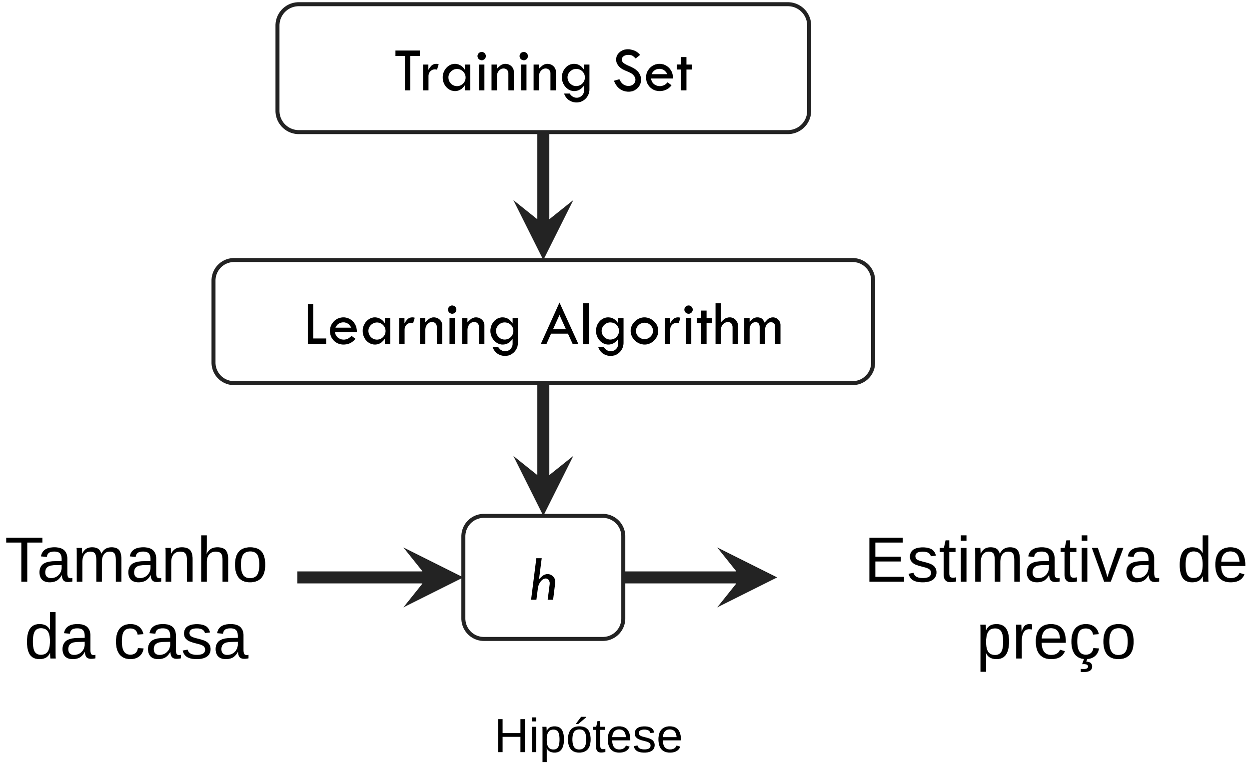
Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Notação:

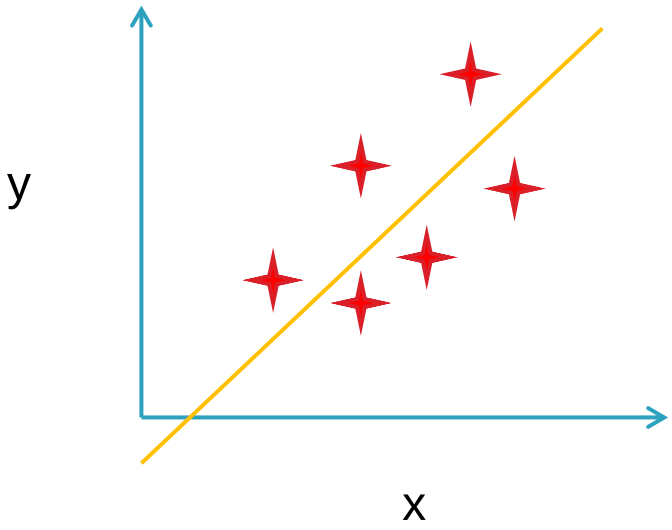
m = Número de exemplos de treinamento

x’s = “input” variable / features

y’s = “output” variable / “target” variable


Training Set
Learning Algorithm
Tamanho da casa
h
Estimativa de preço
Hipótese

Como representar o *h* ?



$$h_\theta(x) = \theta_0 + \theta_1 x$$

Regressão linear com uma variável.
(Univariate linear regression)

Regressão linear com uma variável – Função Custo

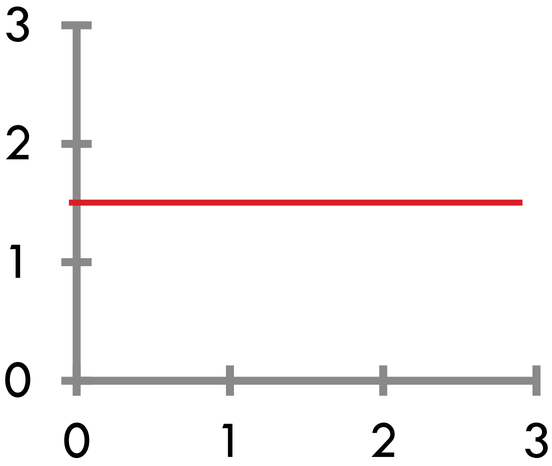
Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Hipótese: $$h_\theta(x) = \theta_0 + \theta_1 x$$

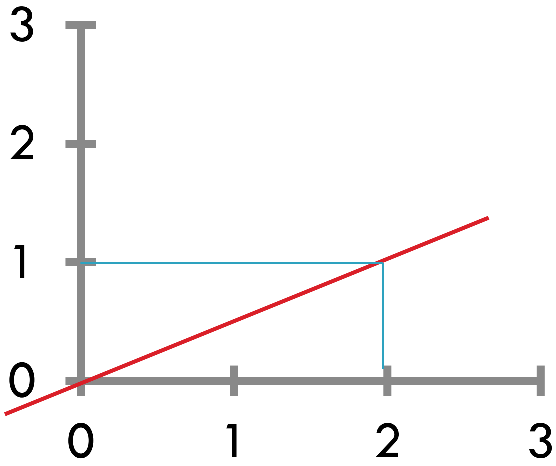
θ_i : Parâmetros

Como escolher θ_i?

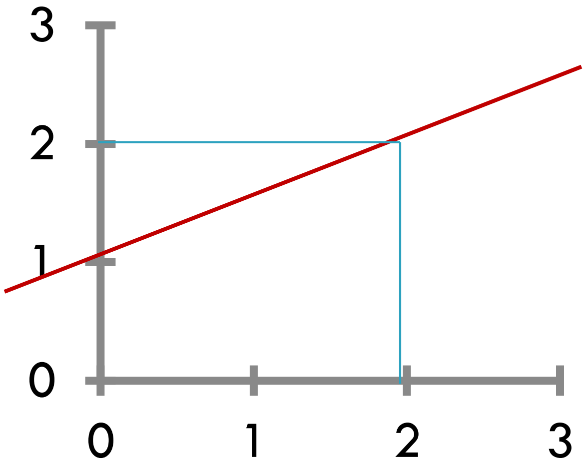
$$h_\theta(x) = \theta_0 + \theta_1 x$$



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

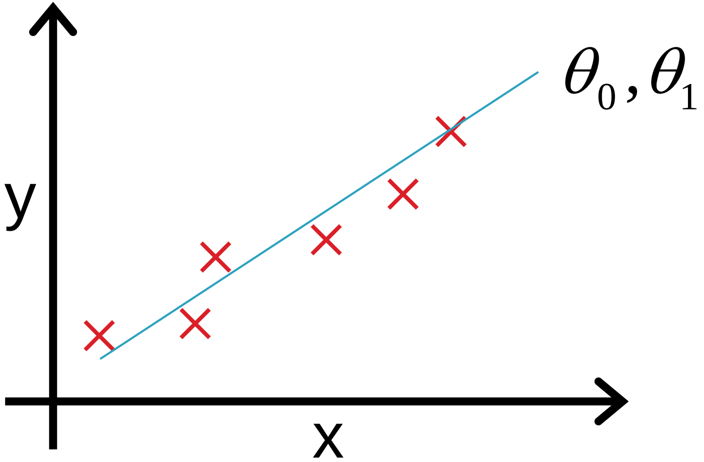


$$\theta_0 = 0$$
$$\theta_1 = 0.5$$



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

Função Custo – Erro médio quadrático



$$\min imize \quad J(\theta_0, \theta_1)$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

sendo

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

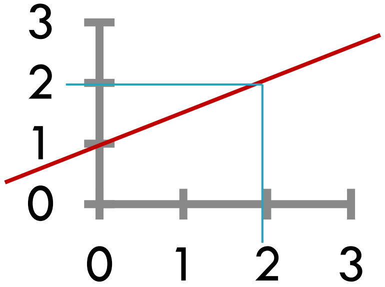
Ideia: escolher θ_0, θ_1, tal que

$h_\theta(x)$ esteja próximo de y para

o conjunto de treinamento (x, y)

Hipótese

$$h_\theta(x) = \theta_0 + \theta_1 x$$



Parâmetros:

$$\theta_0, \theta_1$$

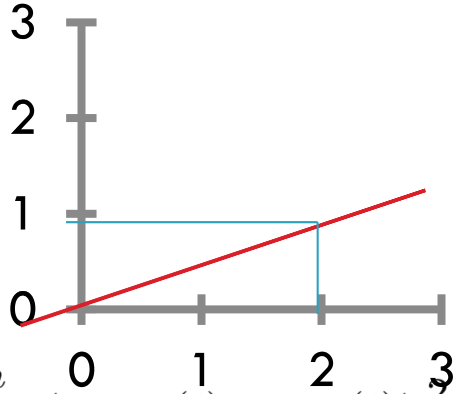
Função Custo:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Objetivo: $$\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$$

Simplificado

$$h_\theta(x) = \theta_1 x$$



$$\theta_1$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

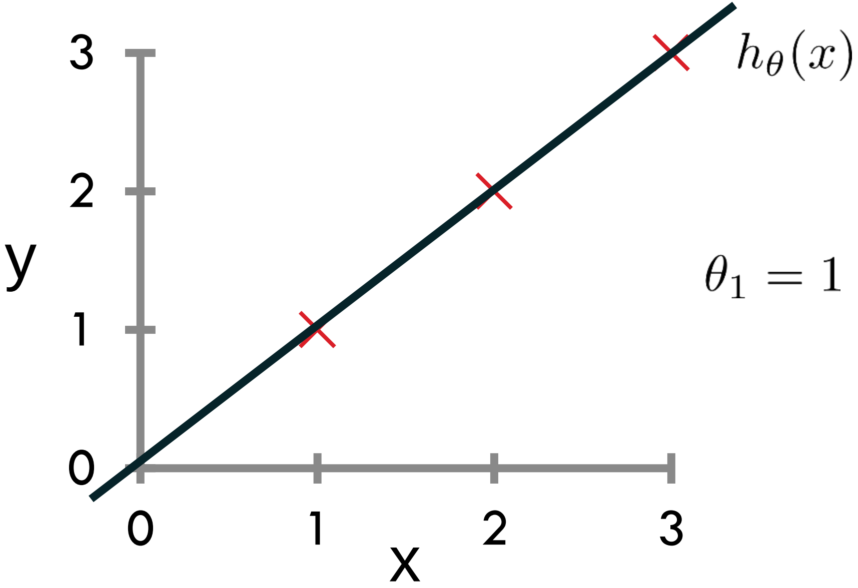
$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

$h_\theta(x)$

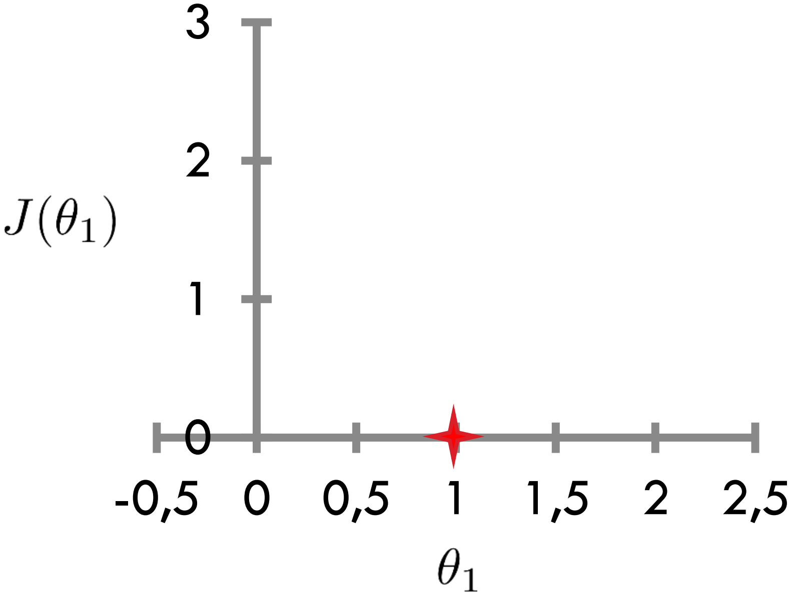
para θ_1 fixo

$J(\theta_1)$

(função do parâmetro θ_1)



$\theta_1 = 1$



$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

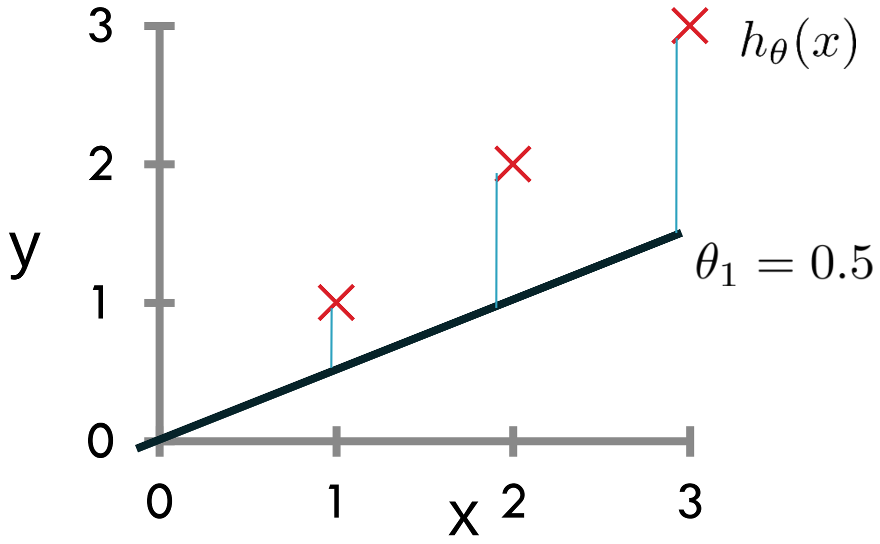
$$= \frac{1}{2m}\sum_{i=1}^{m}\left(\theta_1(x^{(i)}) - y^{(i)}\right)^2 = (0^2 + 0^2 + 0^2) = 0^2$$

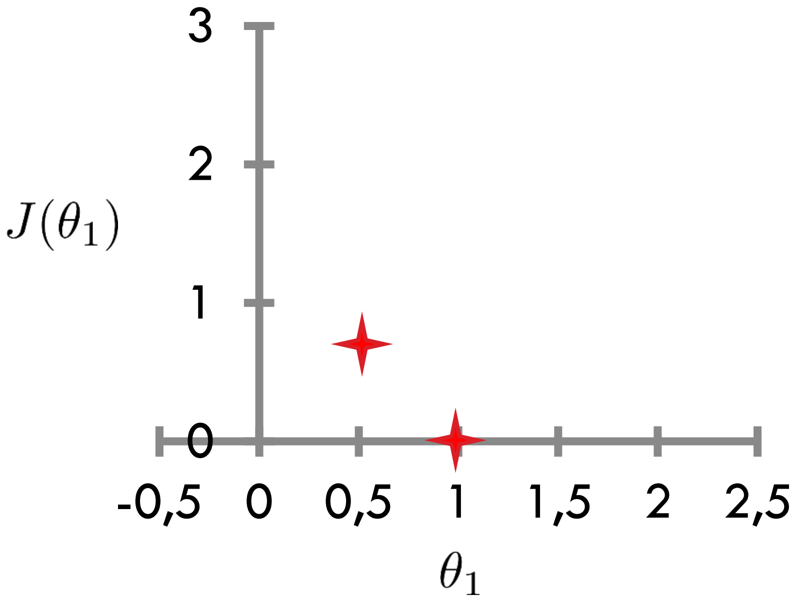
$h_\theta(x)$

para θ_1 fixo

$J(\theta_1)$

(função do parâmetro θ_1)





$$J(0{,}5) = \frac{1}{2m}\left[(0{,}5-1)^2 + (1-2)^2 + (1{,}5-3)^2\right]$$

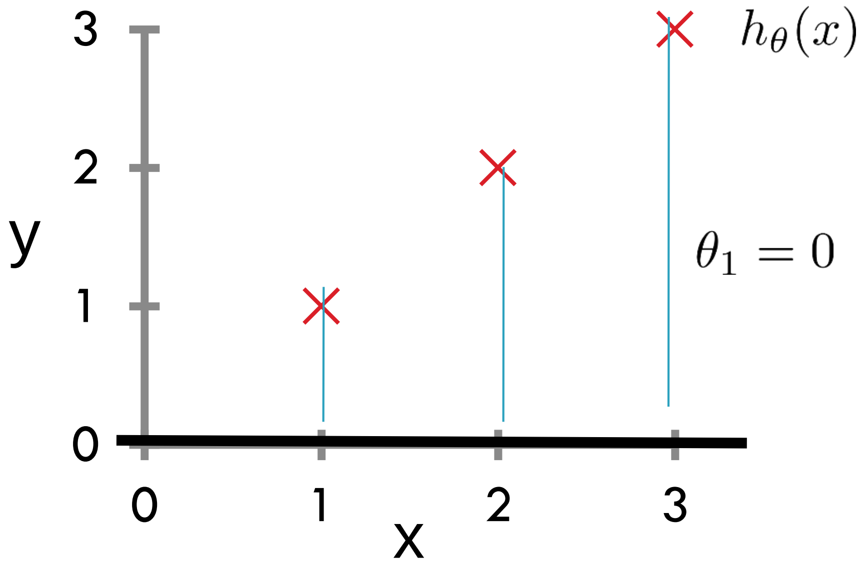
$$= \frac{1}{2\times 3}(3{,}5) = \frac{3{,}5}{6} \approx 0{,}68$$

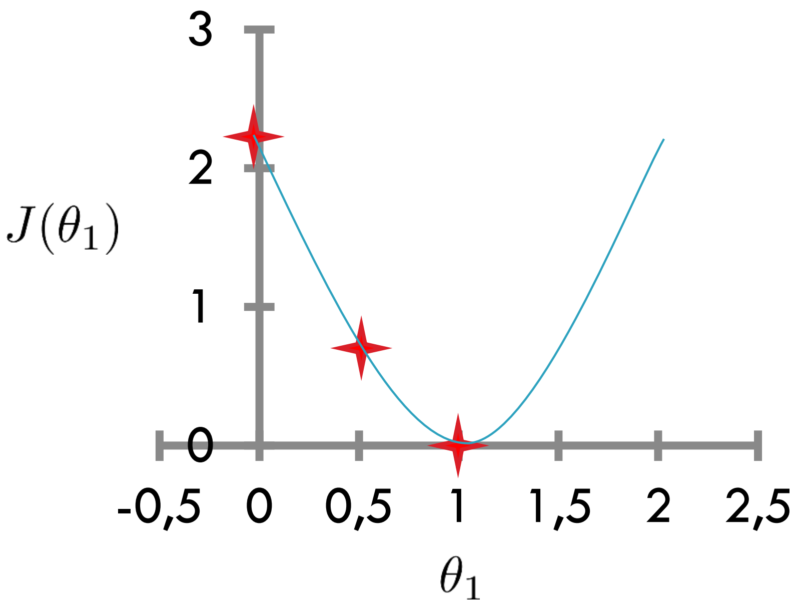
$h_\theta(x)$

para θ_1 fixo

$J(\theta_1)$

(função do parâmetro θ_1)





$$J(0) = \frac{1}{2m}\left[1^2 + 2^2 + 3^2\right]$$

$$= \frac{1}{2 \times 3}(14) \approx 2{,}3$$

Hipótese: $h_\theta(x) = \theta_0 + \theta_1 x$

Parâmetros: θ_0, θ_1

Função Custo: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

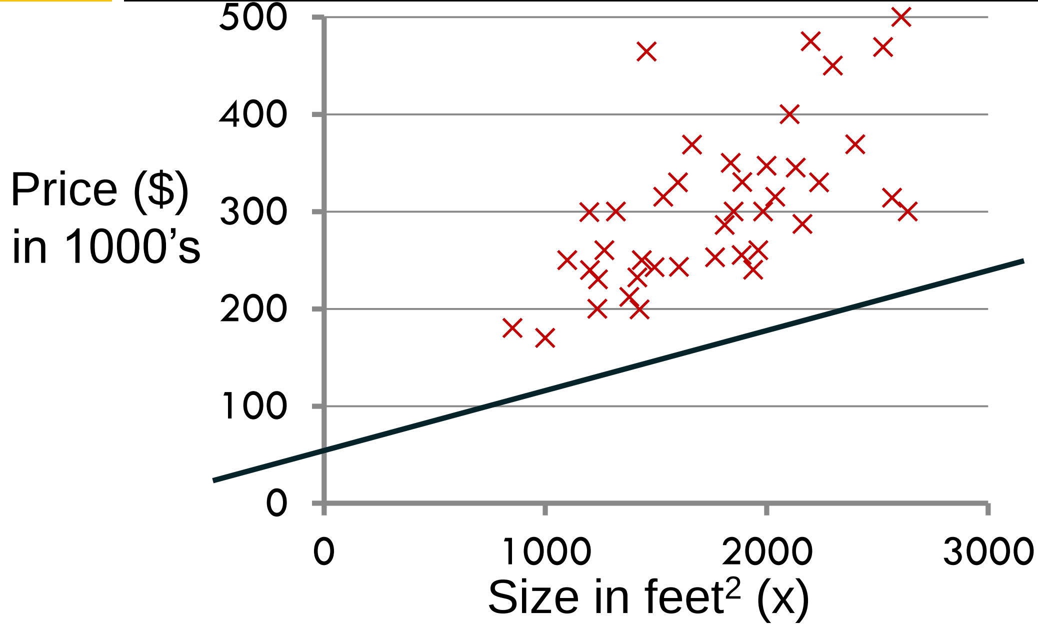
Objetivo: $\underset{\theta_0, \theta_1}{\text{minimize}}\, J(\theta_0, \theta_1)$

$h_\theta(x)$

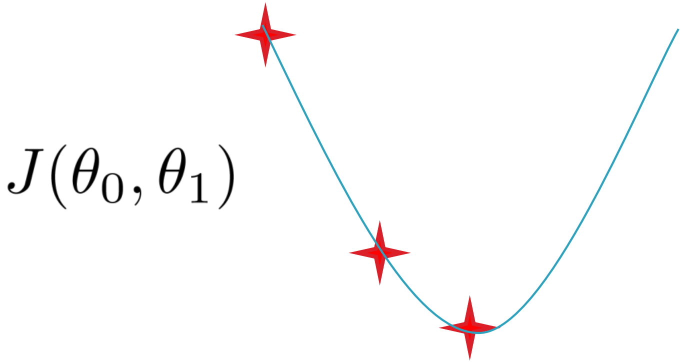
Para θ_0, θ_1 fixos

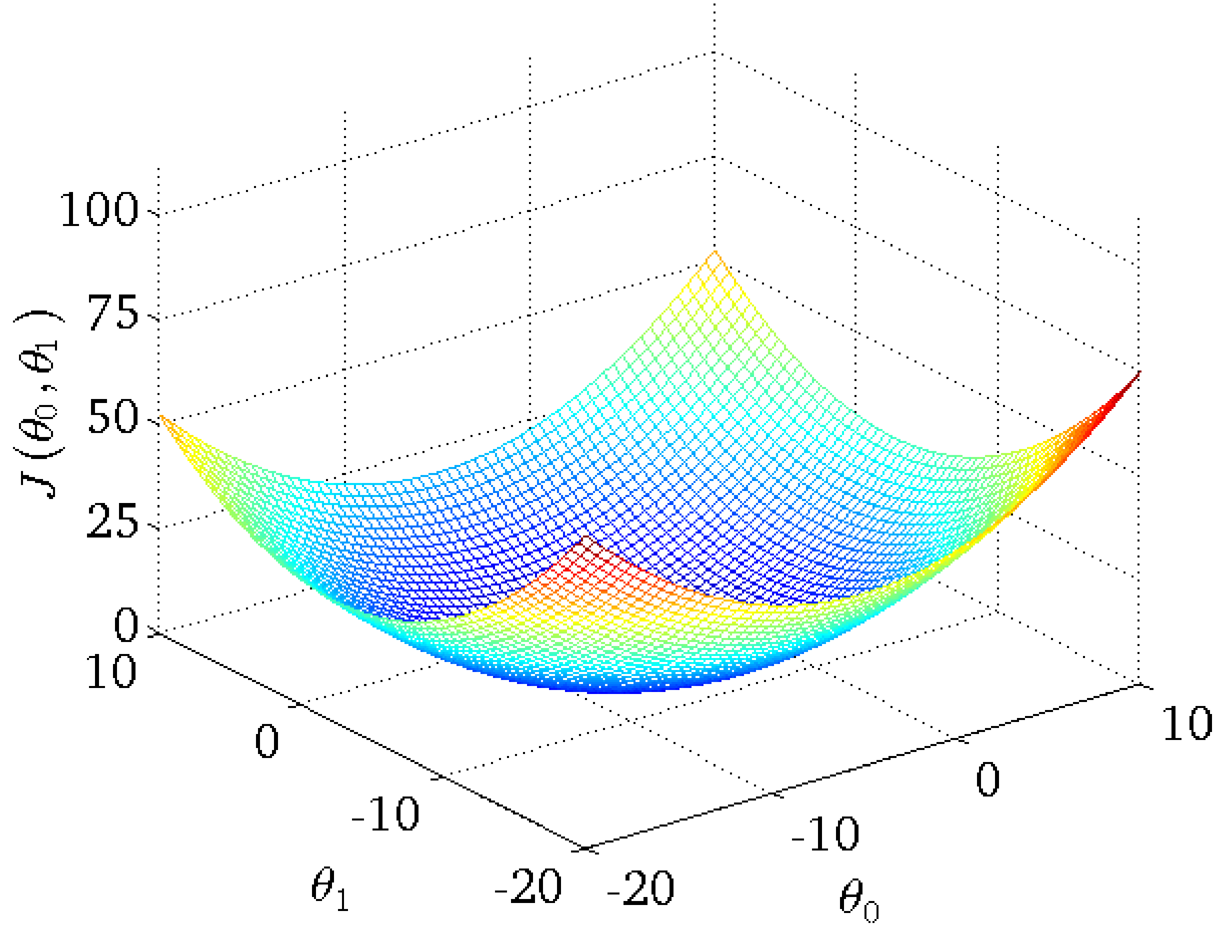
$J(\theta_0, \theta_1)$

(função dos parâmetros θ_0, θ_1)



$h_\theta(x) = 50 + 0.06x$



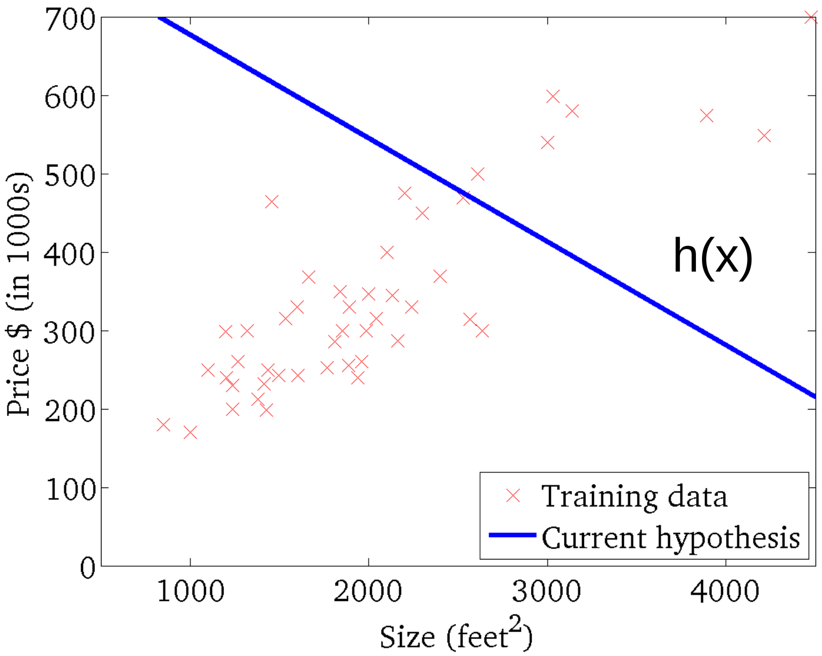

100
75
50
25
0
$J(\theta_0, \theta_1)$
10
0
-10
-20
θ_1
-20
-10
0
10
θ_0

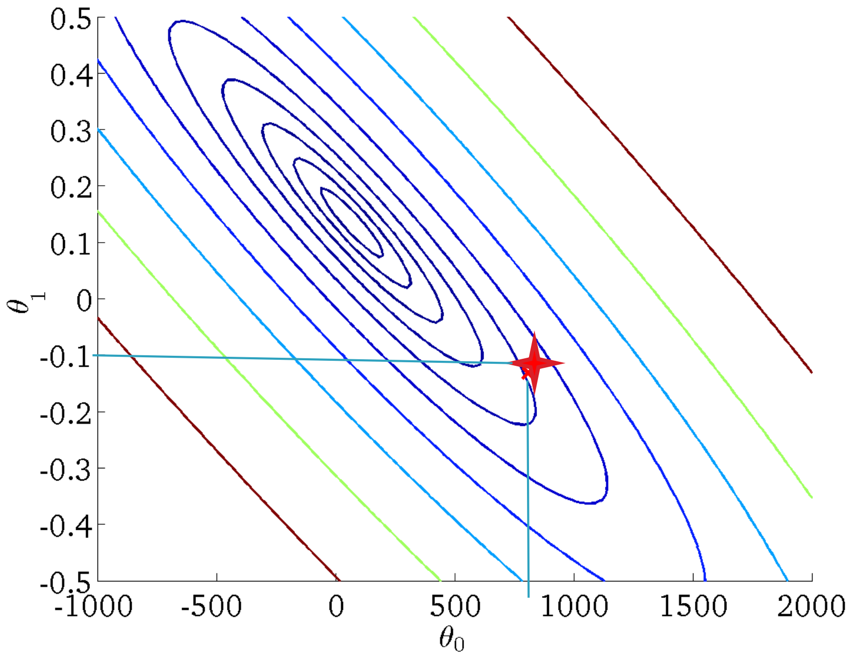
$$h_\theta(x)$$

Para θ_0, θ_1 fixos

$$J(\theta_0, \theta_1)$$

(função dos parâmetros θ_0, θ_1)



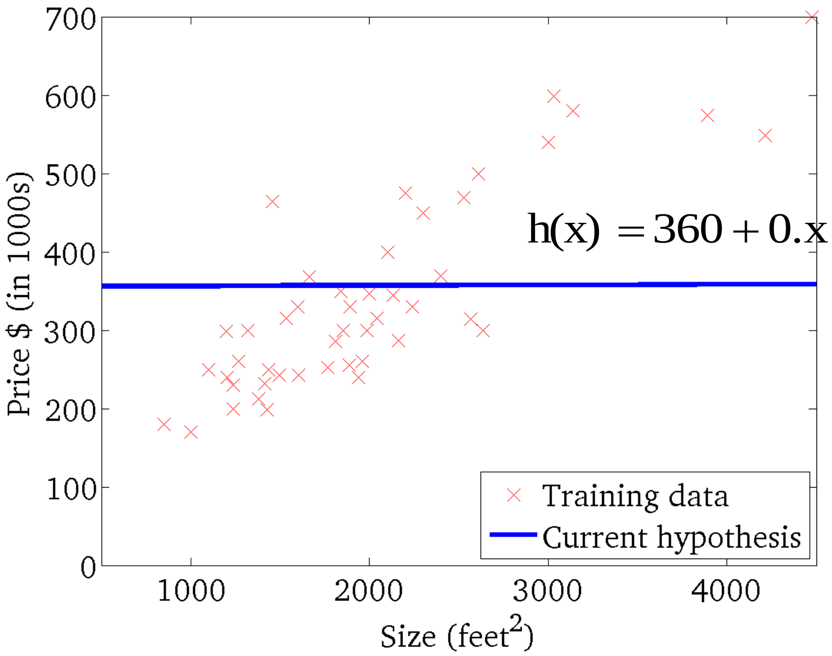


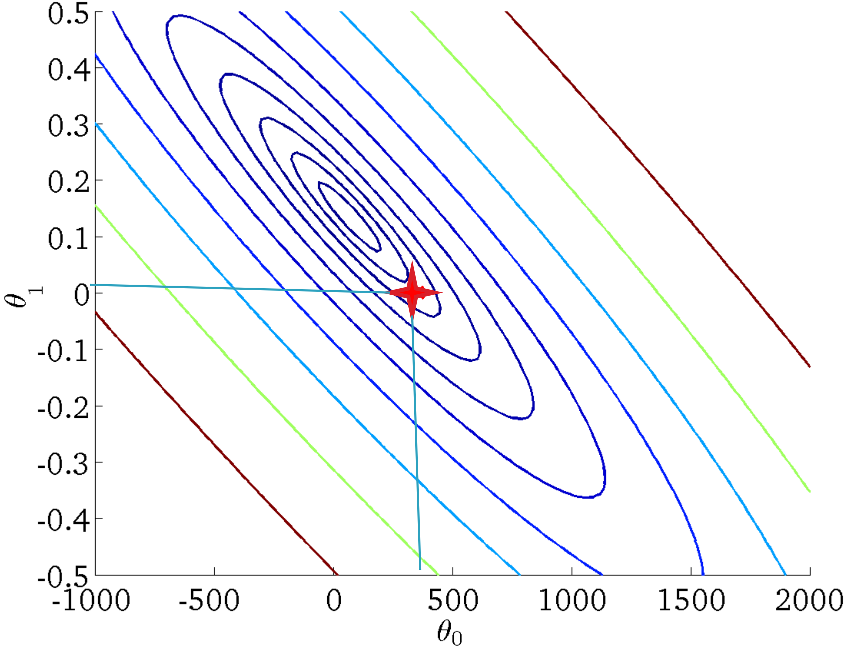
$h_\theta(x)$

Para θ_0, θ_1 fixos

$J(\theta_0, \theta_1)$

(função dos parâmetros θ_0, θ_1)





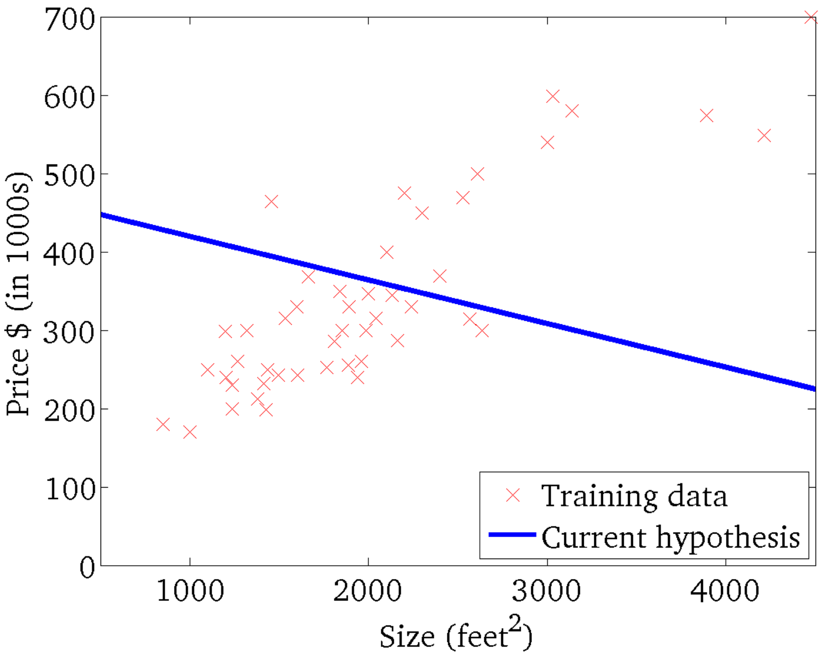
$\theta_0 = 360 \quad \theta_1 = 0$

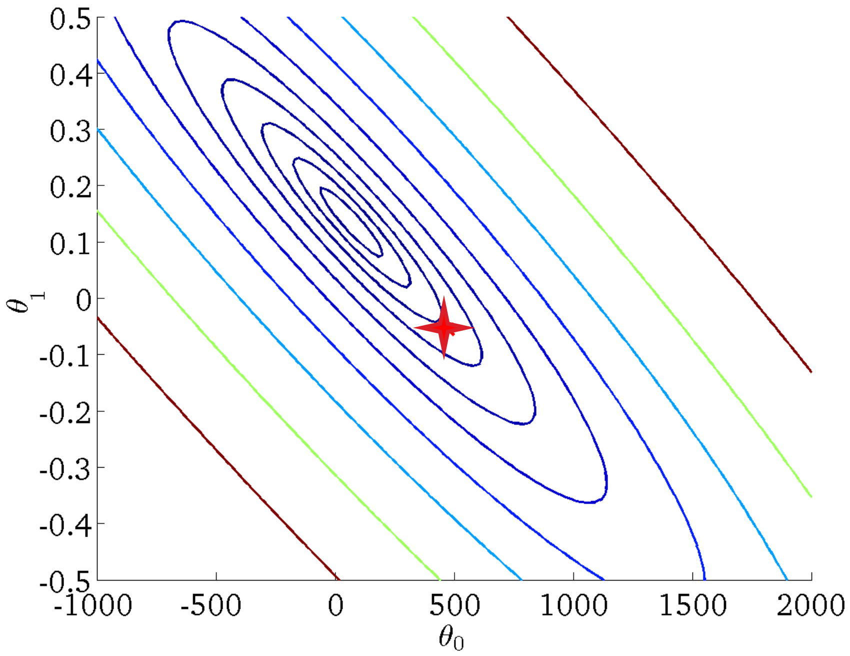
$$h_\theta(x)$$

Para θ_0, θ_1 fixos

$$J(\theta_0, \theta_1)$$

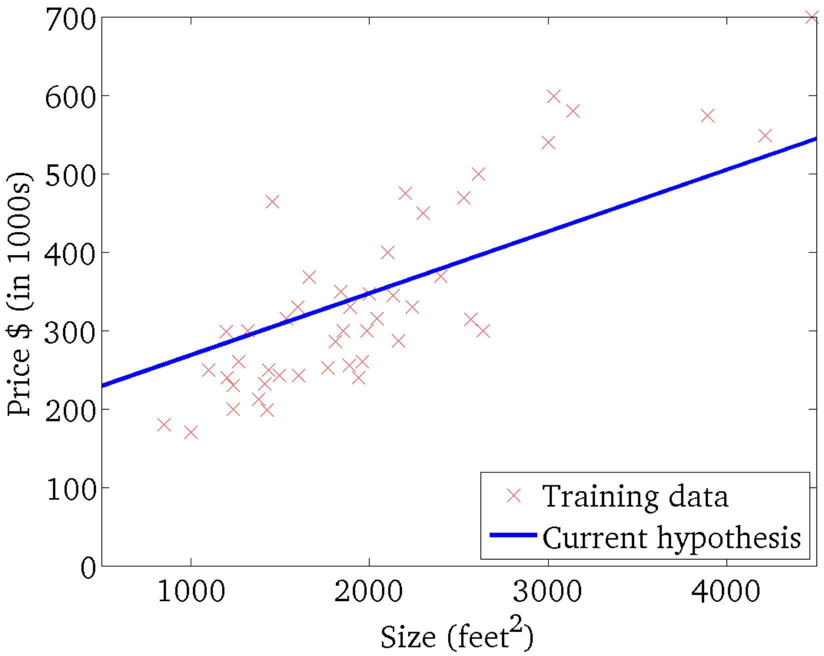
(função dos parâmetros θ_0, θ_1)

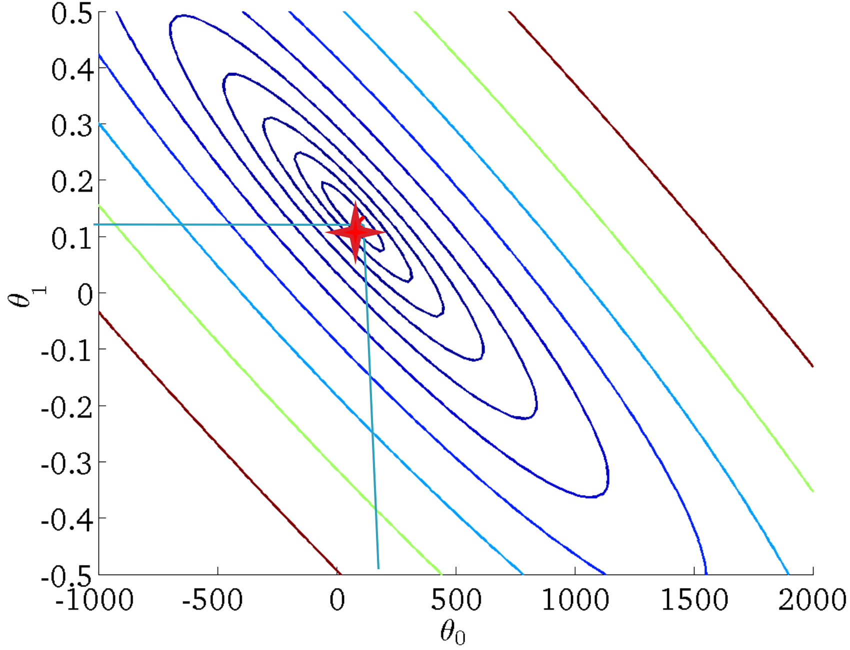




$h_\theta(x)$ $\qquad\qquad$ $J(\theta_0, \theta_1)$

Para θ_0, θ_1 fixos $\qquad$ (função dos parâmetros θ_0, θ_1)





Regressão Linear com uma variável

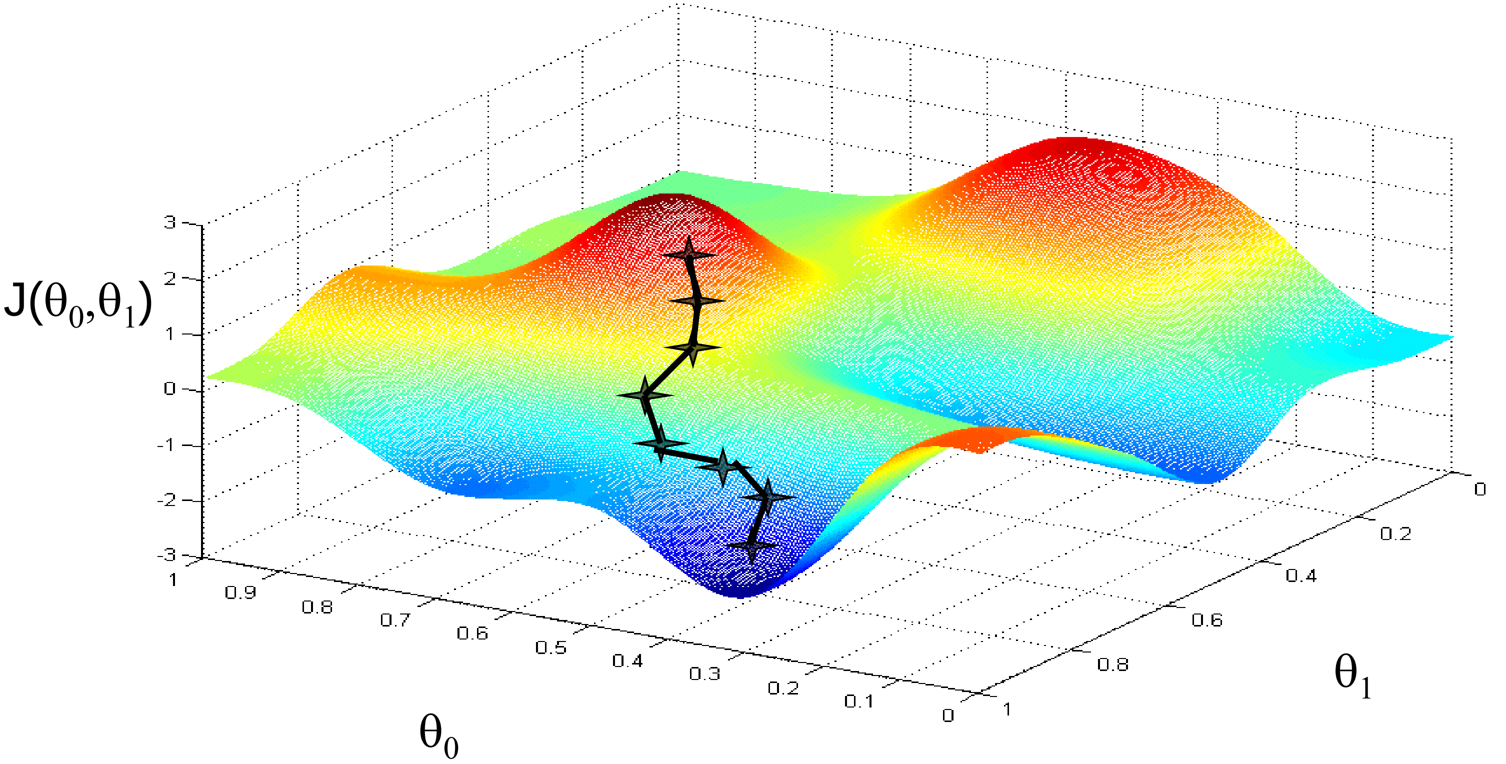
Gradiente descendente

Tem alguma função $J(\theta_0, \theta_1)$

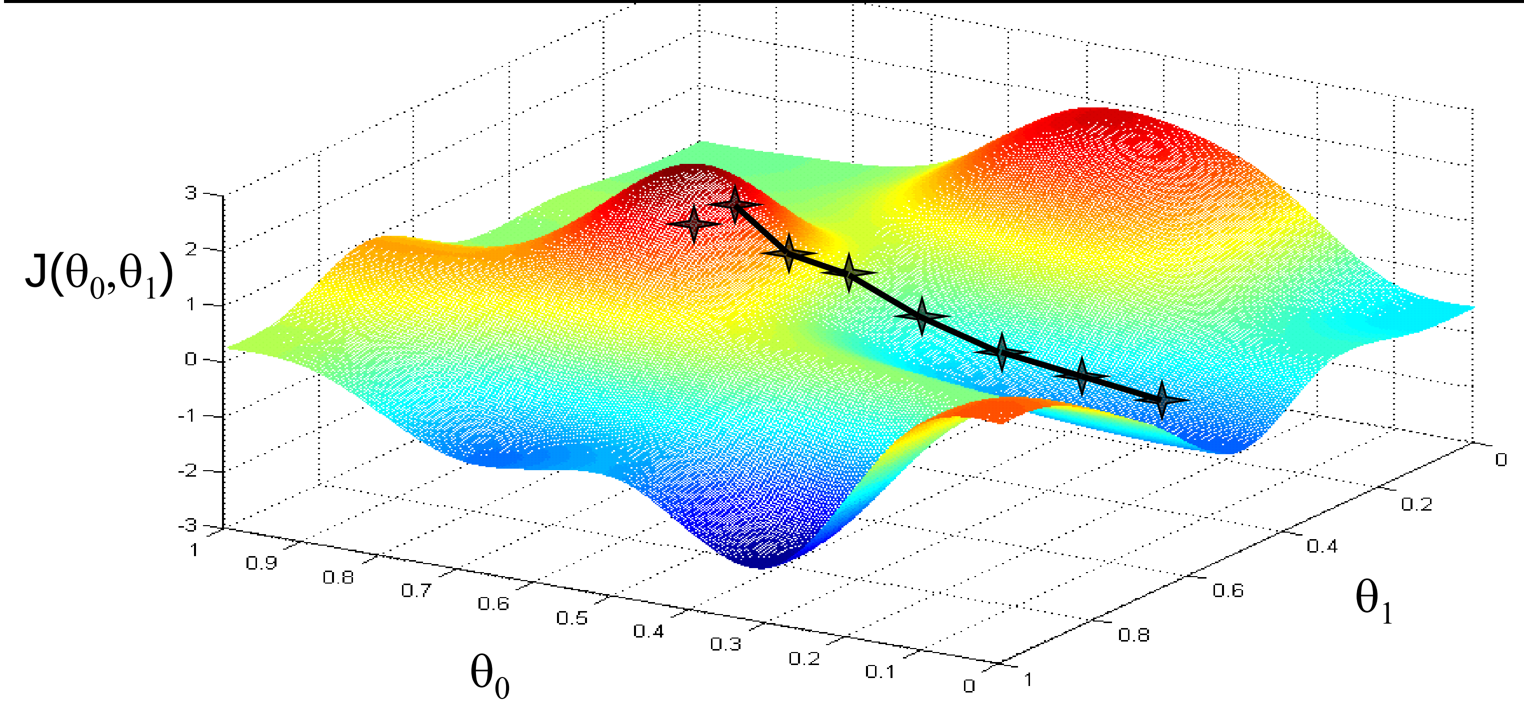
Que $\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$

Outline:

- Comece com algum valor de θ_0, θ_1
- Mude os valores de θ_0, θ_1 de forma a reduzir $J(\theta_0, \theta_1)$ até chegar no valor mínimo



$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1


J(θ0,θ1)
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1

Algoritmo: Gradiente Descendente

$$\text{repeat until convergence } \{ \qquad (\text{for } j = 0 \text{ and } j = 1)$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$
$$\}$$

$\alpha \rightarrow$ *Taxa de apredizagem*

Correto: Atualização Simultânea

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

Incorreto:

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

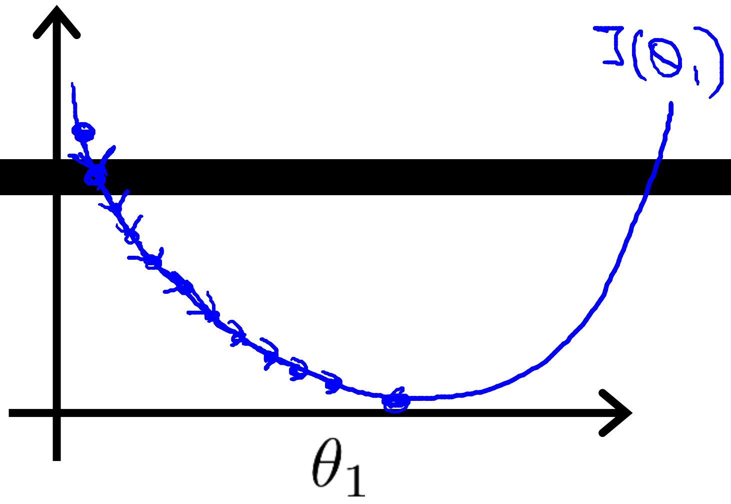
Regressão Linear com uma variável

Algoritmo Gradiente Descendente

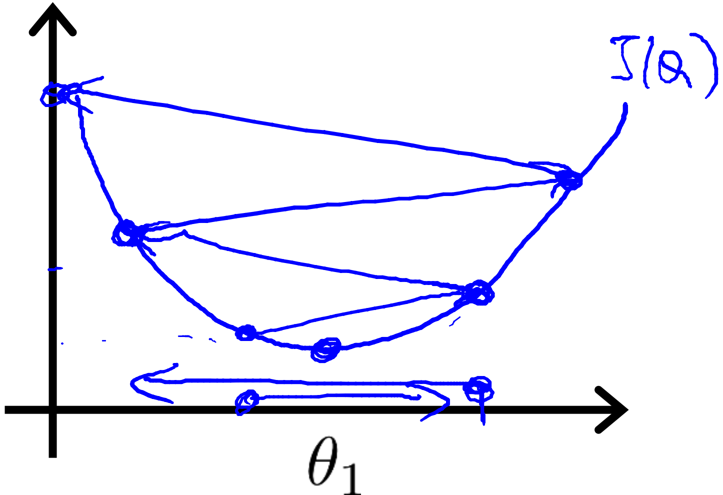
$$
\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad \begin{array}{c}\text{(simultaneously update} \\ j = 0 \text{ and } j = 1)\end{array} \\
&\}
\end{aligned}
$$

$$\underset{\theta_1}{\text{minimize}}\; J(\theta_1)$$

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

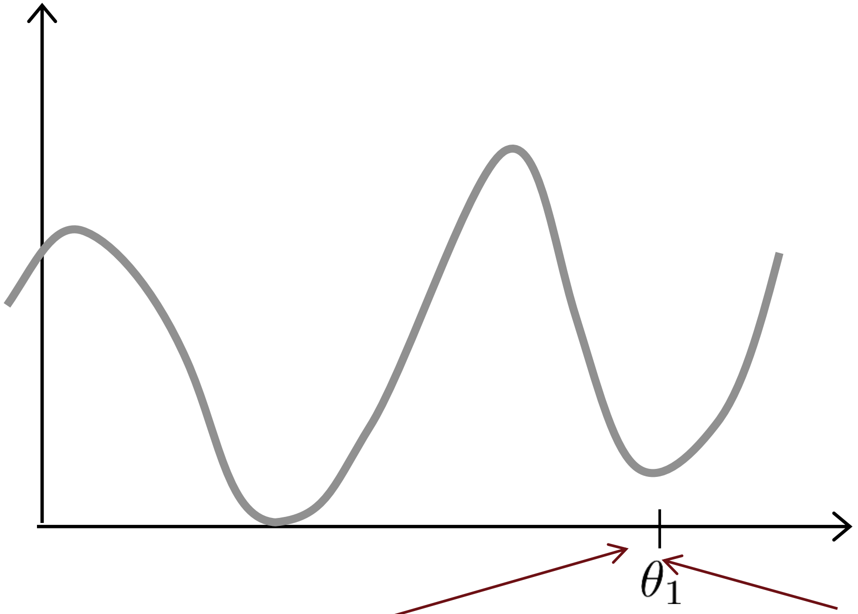


Se α é muito pequeno, o gradiente descendente pode ser muito lento.



Se α é muito grande, o gradiente descendente pode ultrapassar o mínimo. Ele, também, poderia não convergir ou, até mesmo, divergir.

O que acontece se inicializarmos em um mínimo local



θ_1 no ótimo local

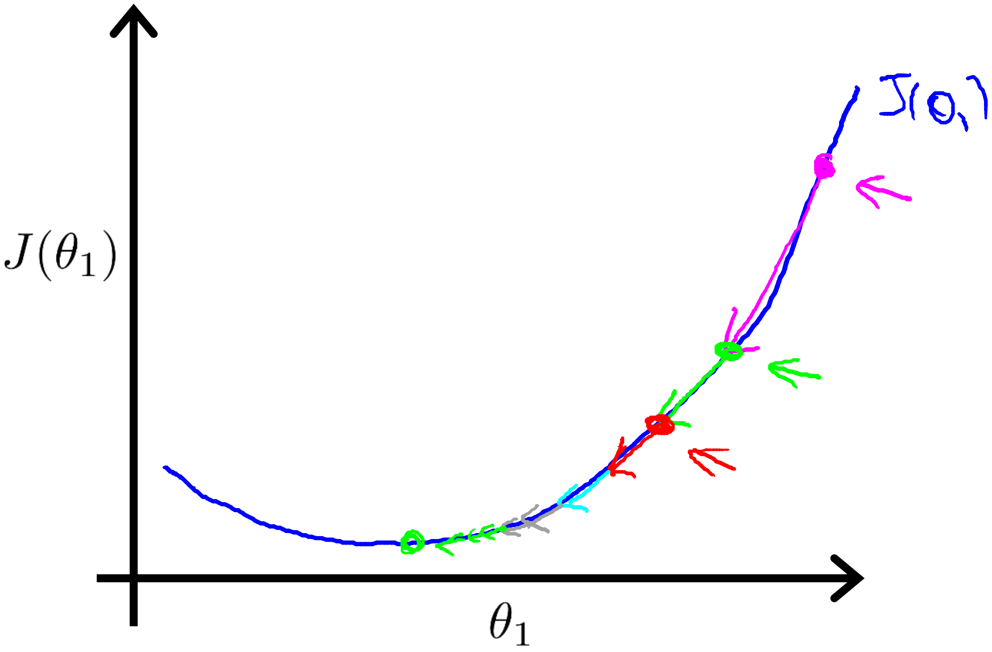
Valor corrente de θ_1

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

Gradiente descendente pode convergir para um mínimo local mesmo com uma taxa de aprendizado (α) fixa

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

Quando se aproxima do mínimo local, o gradiente descendente automaticamente conduzirá a steps menores. Assim, não é necessário diminuir o valor de α ao longo do tempo.



Regressão Linear com uma variável

Gradiente descendente para regressão linear

Algoritmo Gradiente descendente

$$\text{repeat until convergence } \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$
$$(\text{for } j = 1 \text{ and } j = 0)$$
$$\}$$

Modelo de regressão linear

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\frac{\partial}{\partial\theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2$$

$$= \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2$$

$$j = 0 : \frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)$$

$$j = 1 : \frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x^{(i)}$$

Algoritmo Gradiente descendente

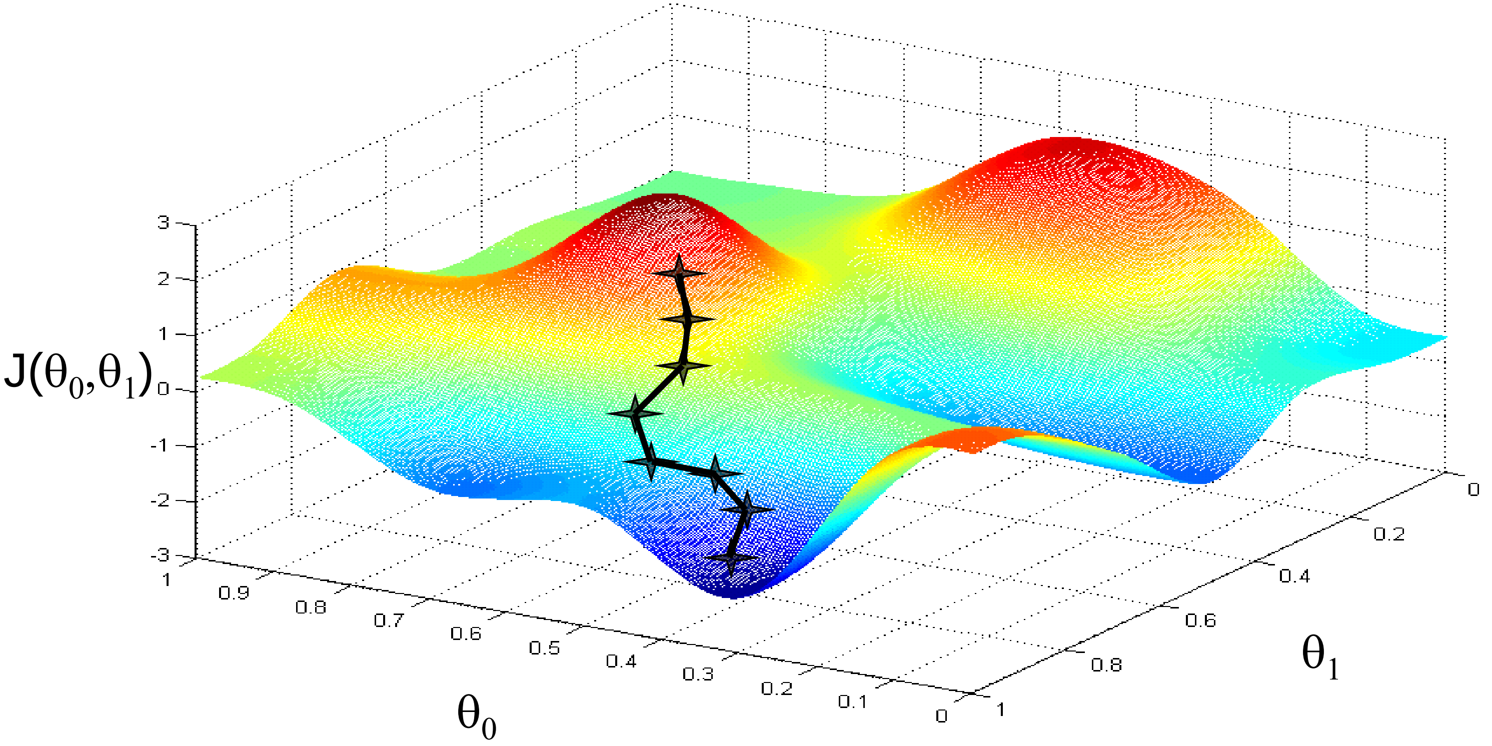
repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

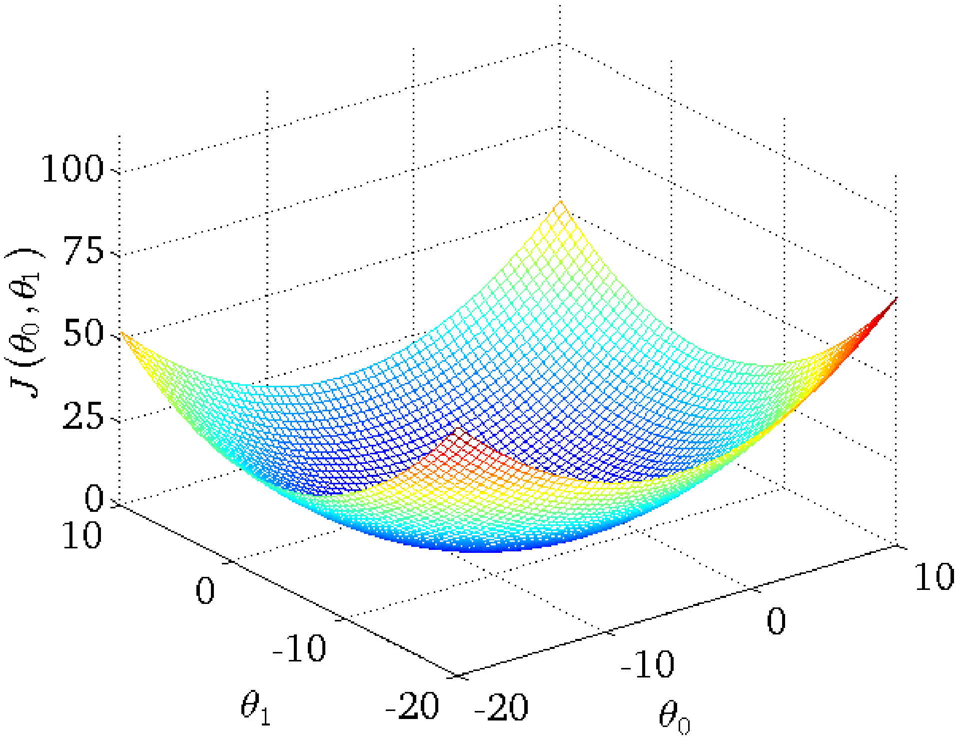
$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}$$

}

Atualize
θ_0 e θ_1
simultaneamente


$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1

Função Convexa

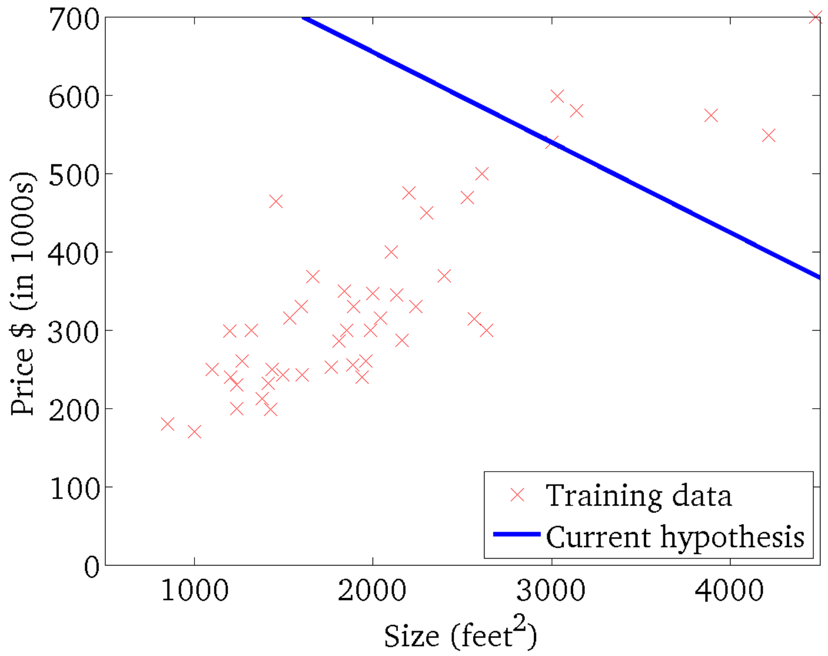


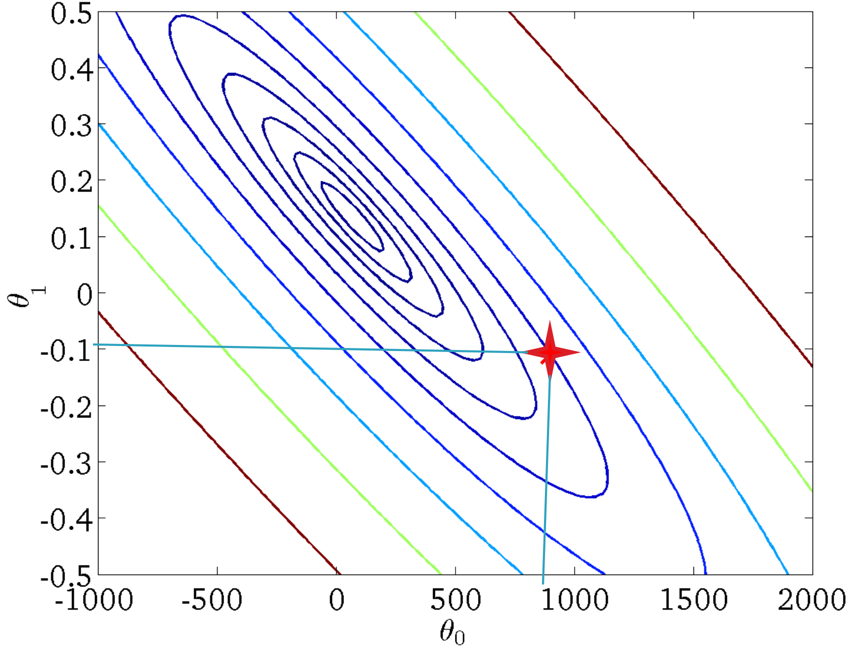
$h_\theta(x)$

(para θ_0, θ_1 fixos)

$J(\theta_0, \theta_1)$

(função custo em função de θ_0, θ_1)



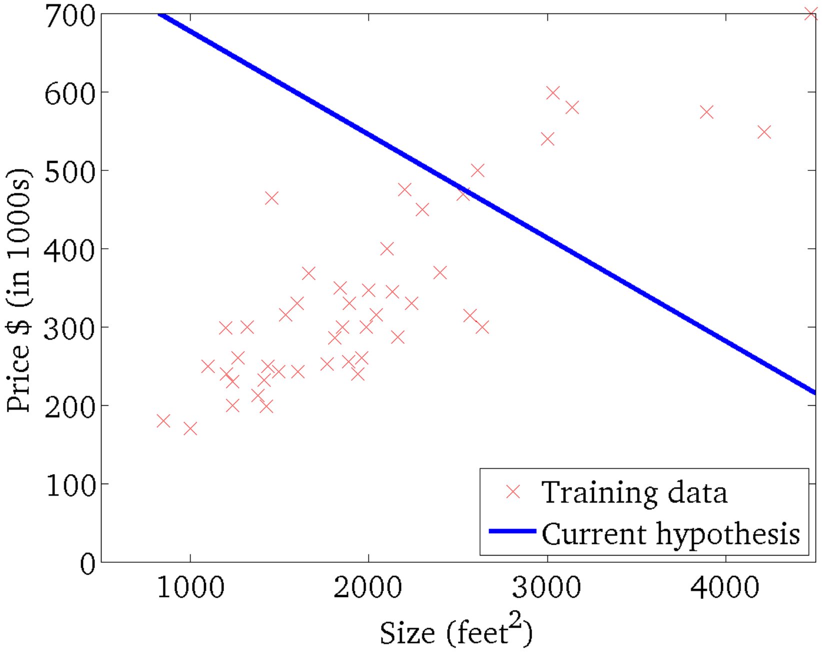


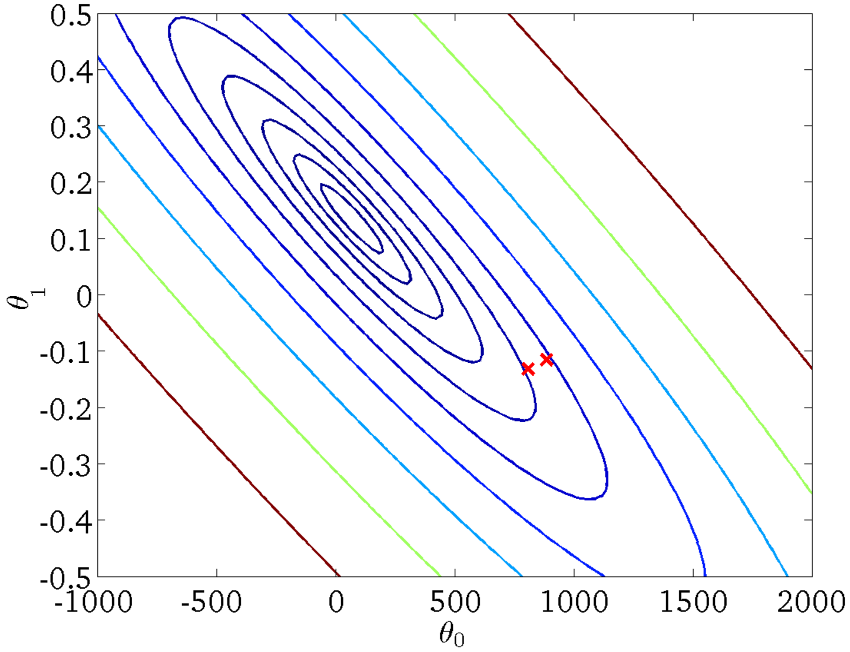
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



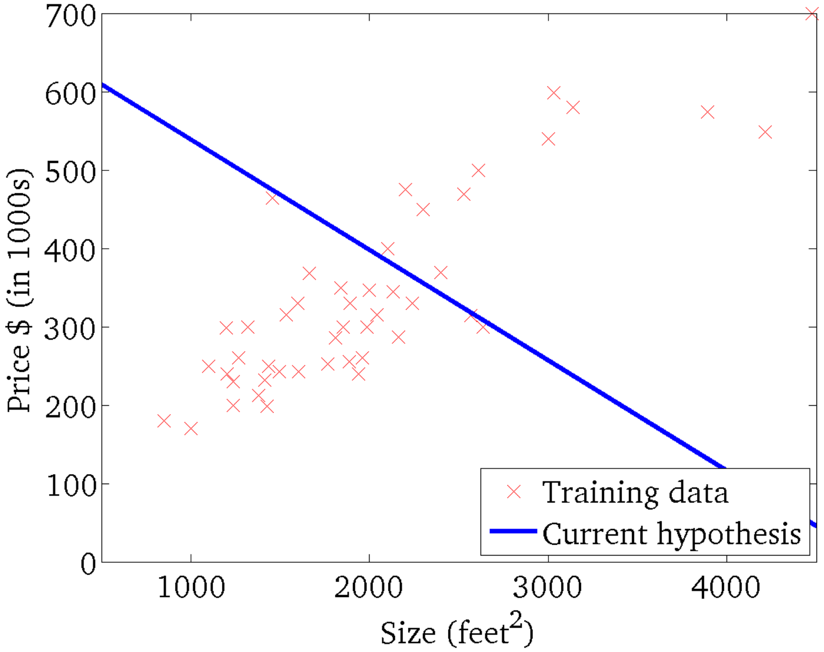


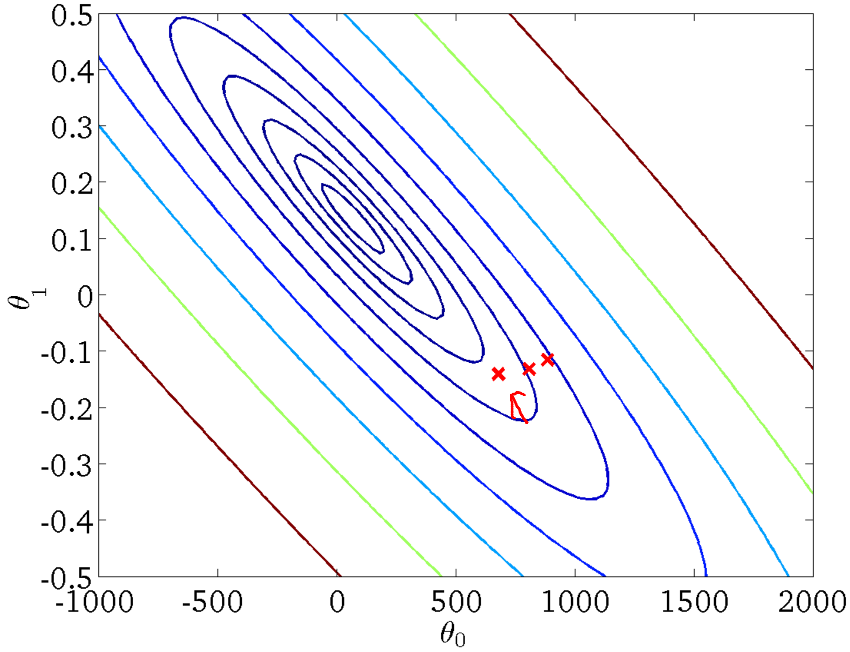
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



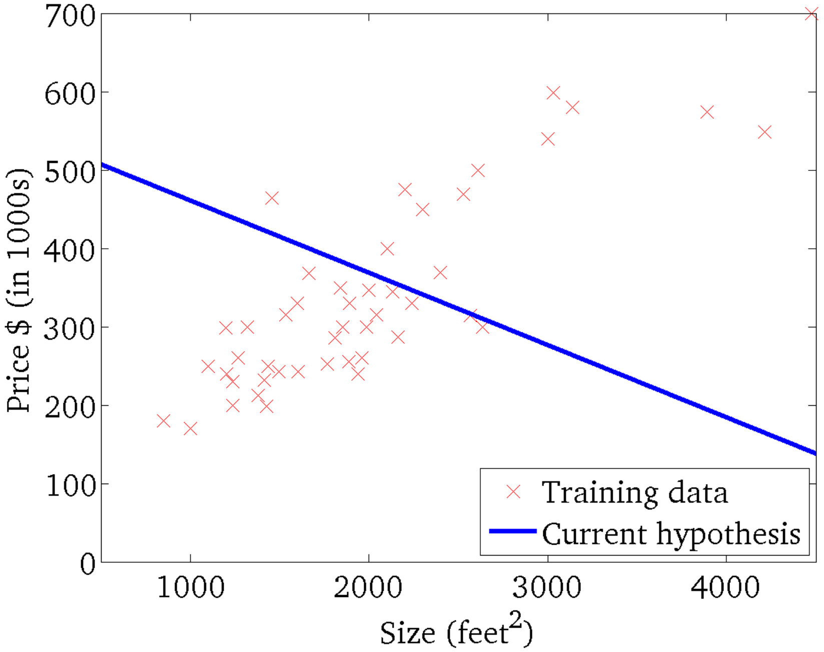


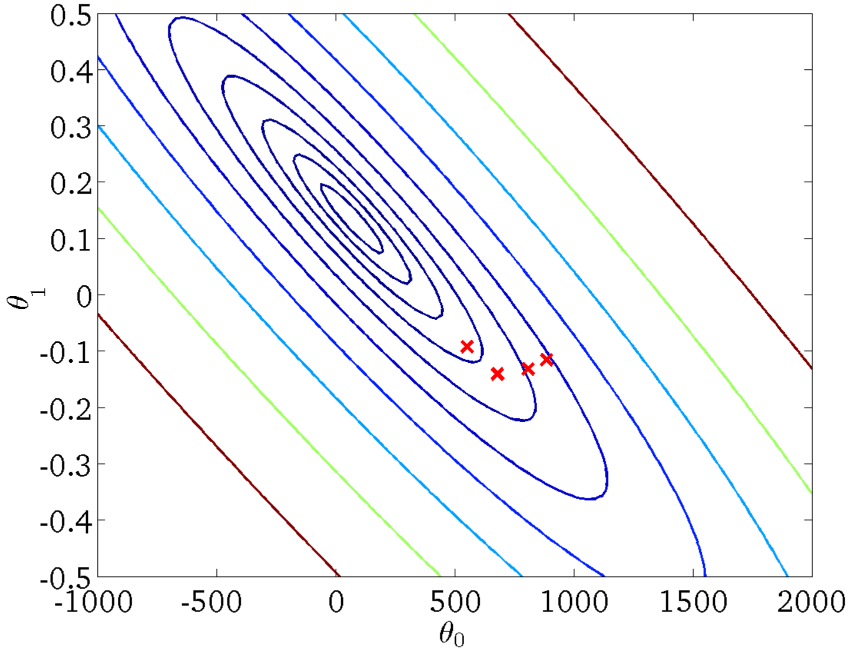
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



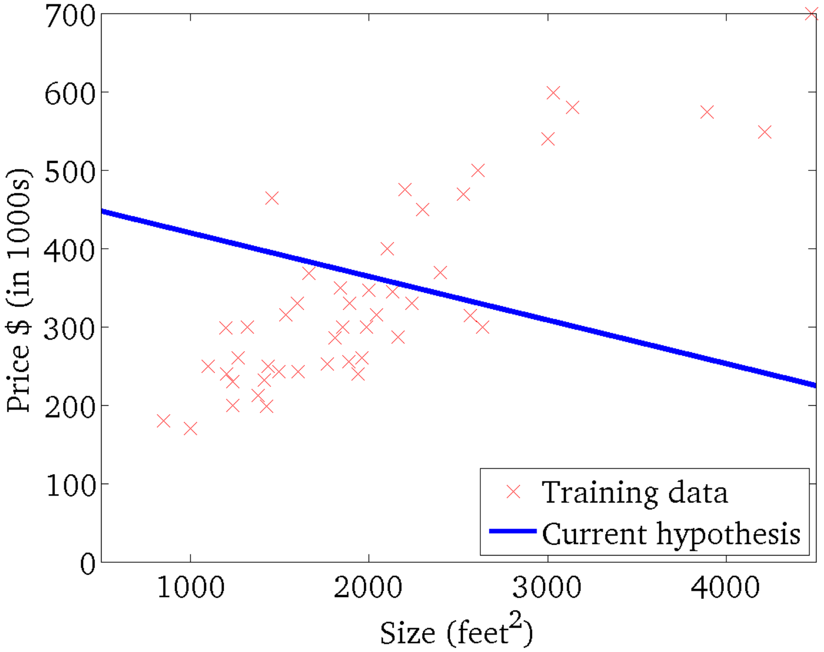


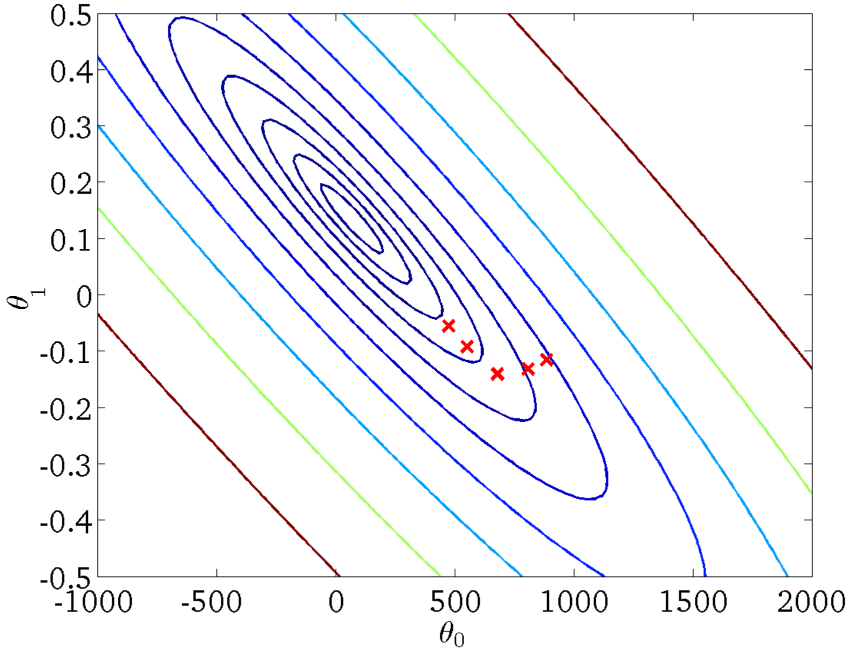
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



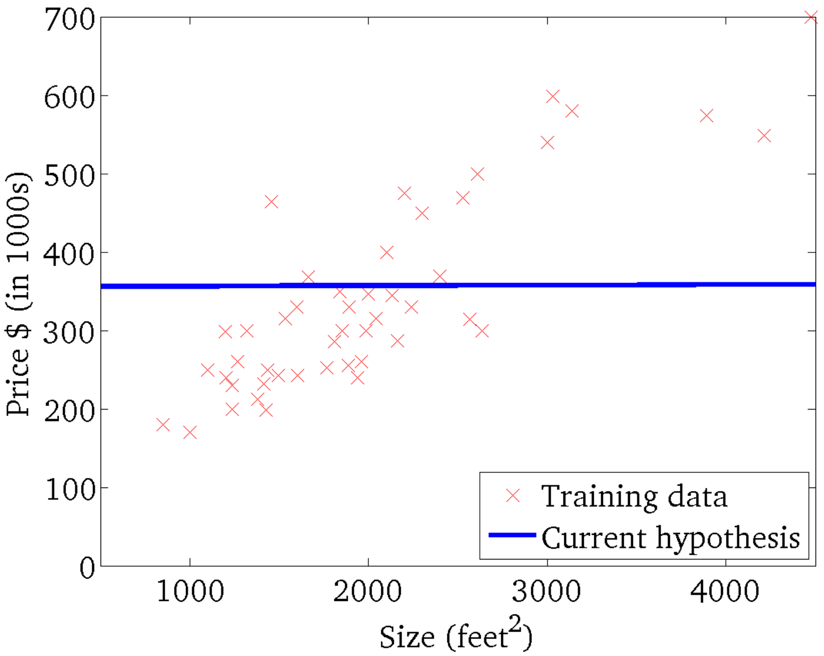


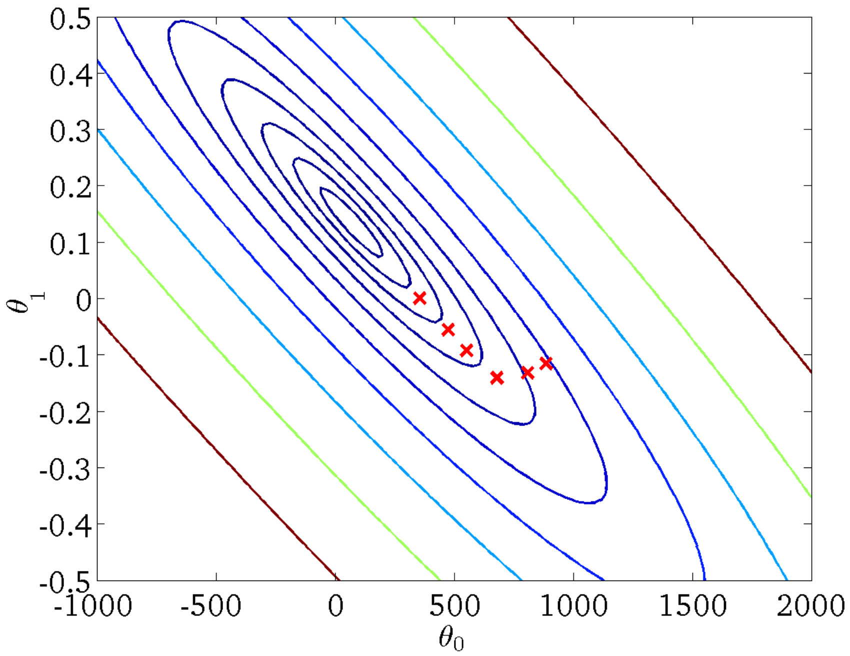
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



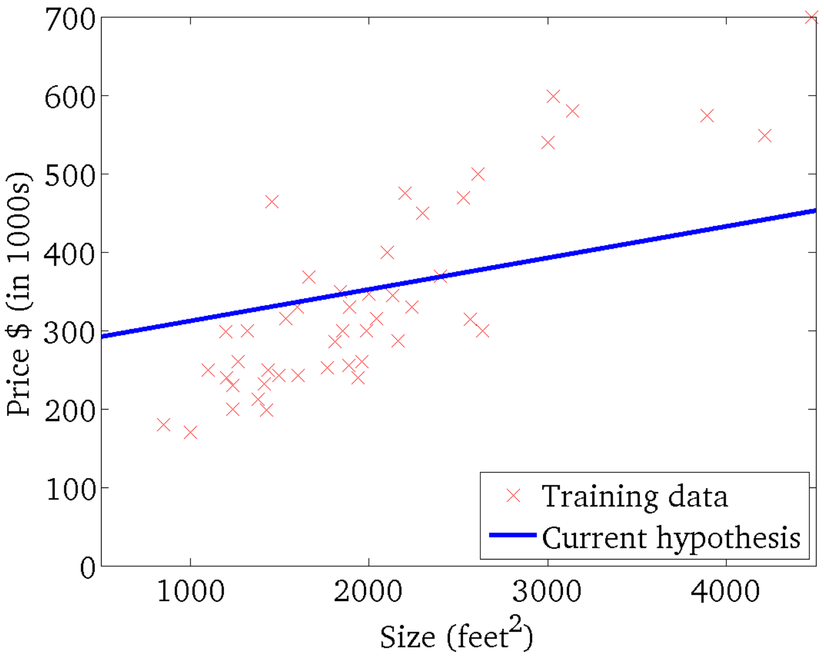


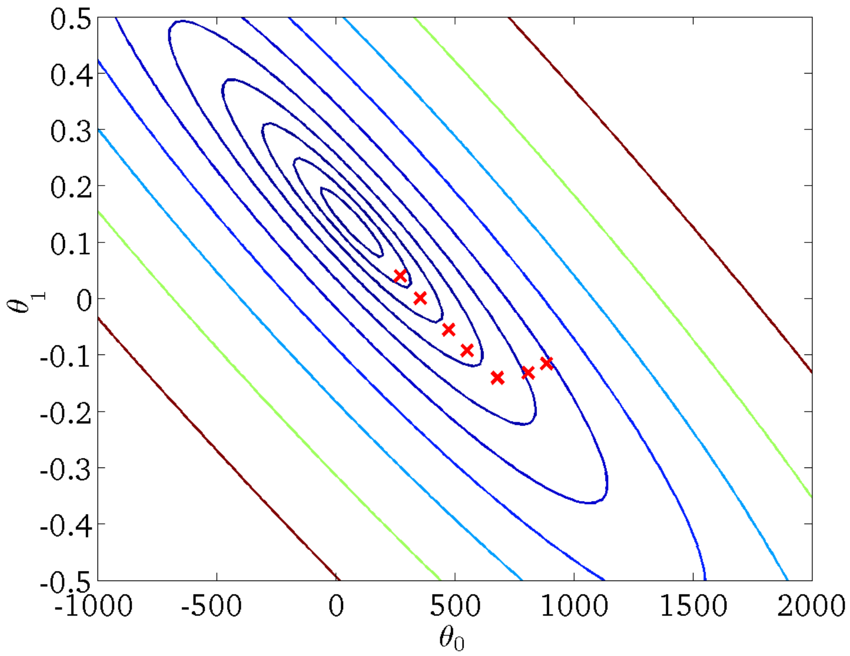
$h_\theta(x)$

(para θ_0, θ_1 fixos)

$J(\theta_0, \theta_1)$

(função custo em função de θ_0, θ_1)



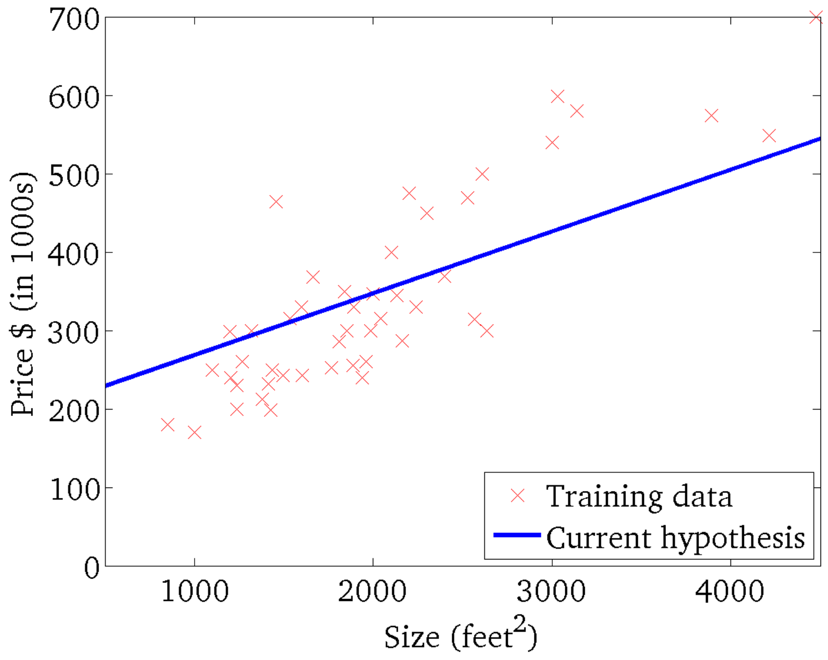


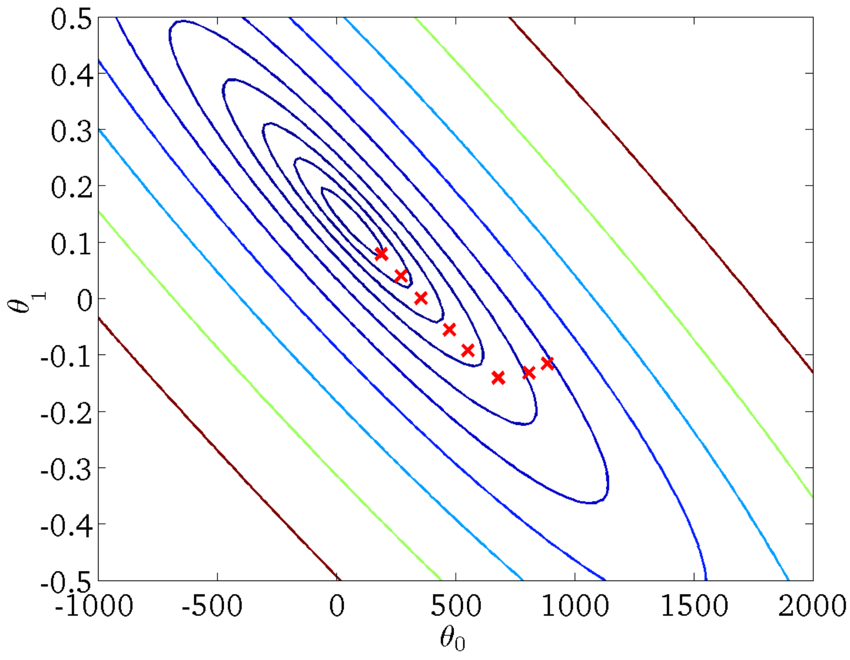
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)



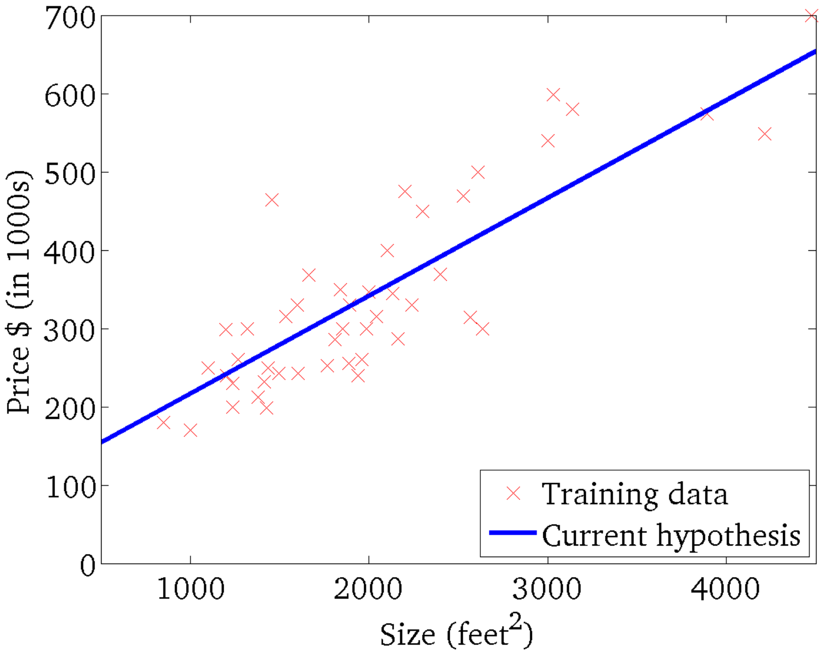


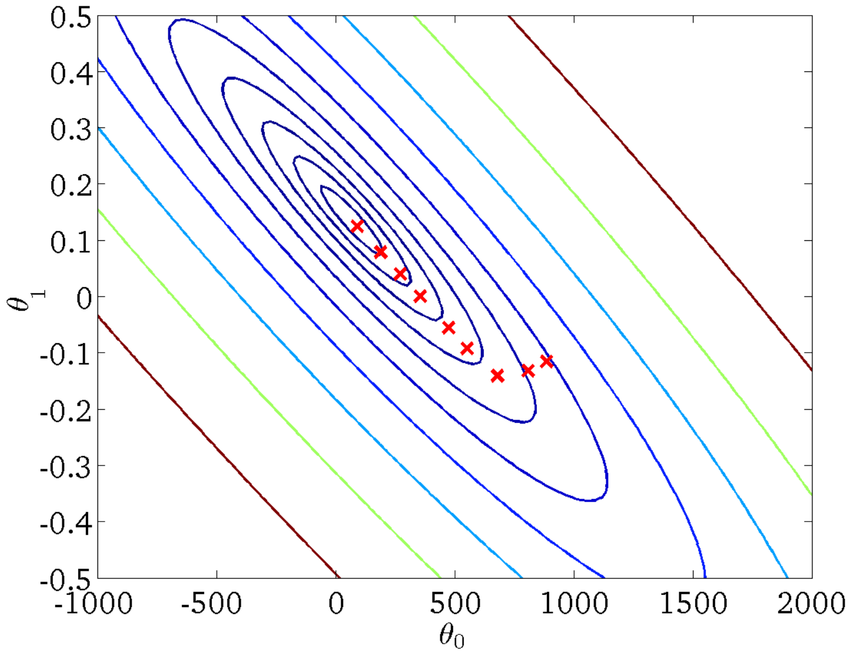
$$h_\theta(x)$$

(para θ_0, θ_1 fixos)

$$J(\theta_0, \theta_1)$$

(função custo em função de θ_0, θ_1)





Gradiente Descente – Treinamento “Batch”

“Batch”: Em cada passo ou época do gradiente descendente use todos os exemplos de treinamento