

Aula 2 - Regressão Linear

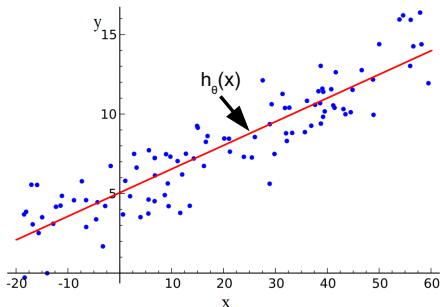
João Florindo

Instituto de Matemática, Estatística e Computação Científica
Universidade Estadual de Campinas - Brasil
florindo@unicamp.br

Outline

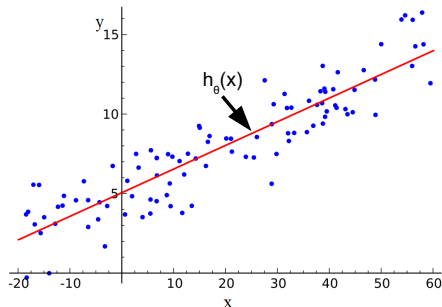
- 1 Regressão Linear
- 2 Gradiente Descendente
- 3 Gradiente Descendente na Regressão Linear

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



- Parâmetros θ_0 e θ_1 são **aprendidos** a partir do conjunto de treinamento $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$.

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



- Parâmetros θ_0 e θ_1 são **aprendidos** a partir do conjunto de treinamento $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$.

Como obter θ_0 e θ_1 ? R.: Minimizar uma **função de custo** $J(\theta)$

FUNÇÃO DE CUSTO

Uma função de custo $J(\theta)$ é uma medida do erro na aproximação de $y^{(i)}$ por $h_{\theta}(x^{(i)})$.

Na regressão linear, a função de custo é o **erro quadrático médio**:

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

Objetivo: obter θ_0 e θ_1 que minimizem $J(\theta_0, \theta_1)$:

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

Como obter θ_0 e θ_1 ? R.: Minimizar uma **função de custo** $J(\theta)$

FUNÇÃO DE CUSTO

Uma função de custo $J(\theta)$ é uma medida do erro na aproximação de $y^{(i)}$ por $h_{\theta}(x^{(i)})$.

Na regressão linear, a função de custo é o **erro quadrático médio**:

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

Objetivo: obter θ_0 e θ_1 que minimizem $J(\theta_0, \theta_1)$:

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

Como obter θ_0 e θ_1 ? R.: Minimizar uma **função de custo** $J(\theta)$

FUNÇÃO DE CUSTO

Uma função de custo $J(\theta)$ é uma medida do erro na aproximação de $y^{(i)}$ por $h_{\theta}(x^{(i)})$.

Na regressão linear, a função de custo é o **erro quadrático médio**:

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

Objetivo: obter θ_0 e θ_1 que minimizem $J(\theta_0, \theta_1)$:

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

Como obter θ_0 e θ_1 ? R.: Minimizar uma **função de custo** $J(\theta)$

FUNÇÃO DE CUSTO

Uma função de custo $J(\theta)$ é uma medida do erro na aproximação de $y^{(i)}$ por $h_{\theta}(x^{(i)})$.

Na regressão linear, a função de custo é o **erro quadrático médio**:

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

Objetivo: obter θ_0 e θ_1 que minimizem $J(\theta_0, \theta_1)$:

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

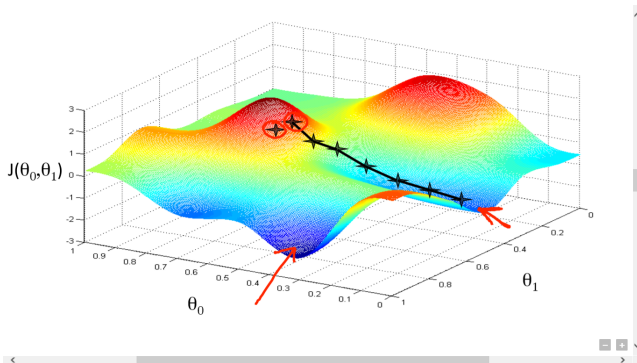
Outline

- 1 Regressão Linear
- 2 Gradiente Descendente
- 3 Gradiente Descendente na Regressão Linear

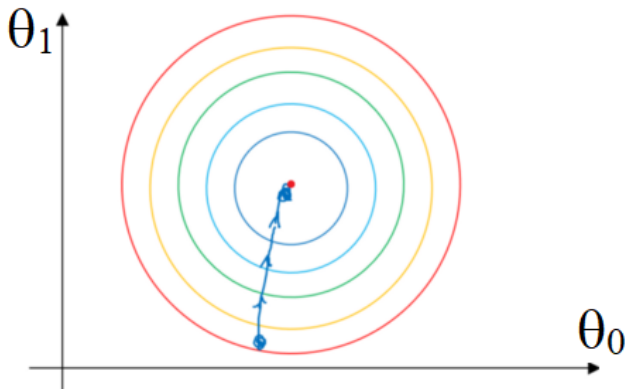
Obter θ que minimize $J(\theta)$:

Repita até convergir:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta),$$



Source:
<https://hackernoon.com/gradient-descent-aynk-7cbe95a778da>



Source: Adaptado de
https://miro.medium.com/max/875/1*vXpnsIMSpOELhovg.png

ATENÇÃO: Atualização SIMULTÂNEA.

CORRETO

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

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

$$\theta_0 := \text{temp0}$$

$$\theta_1 := \text{temp1}$$

INCORRETO

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

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

Outline

- 1 Regressão Linear
- 2 Gradiente Descendente
- 3 Gradiente Descendente na Regressão Linear

Lembrando que na regressão linear temos

$$\begin{cases} h_{\theta}(x) = \theta_0 + \theta_1 x \\ J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \end{cases}$$

e no gradiente descendente fazemos

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

Precisamos calcular

$$\frac{\partial}{\partial \theta_j} J(\theta), \quad j = 0, 1.$$

Lembrando que na regressão linear temos

$$\begin{cases} h_{\theta}(x) = \theta_0 + \theta_1 x \\ J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \end{cases}$$

e no gradiente descendente fazemos

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

Precisamos calcular

$$\frac{\partial}{\partial \theta_j} J(\theta), \quad j = 0, 1.$$

Lembrando que na regressão linear temos

$$\begin{cases} h_{\theta}(x) = \theta_0 + \theta_1 x \\ J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \end{cases}$$

e no gradiente descendente fazemos

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

Precisamos calcular

$$\frac{\partial}{\partial \theta_j} J(\theta), \quad j = 0, 1.$$

Assim chegamos ao algoritmo seguinte:

Repita até convergir: {

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

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

}

OBSERVAÇÕES:

- Lembre-se de que a atualização de θ_0 e θ_1 é **SIMULTÂNEA**
- $J(\theta)$ na regressão linear é **convexa**
- Usa todos os $\{x^{(i)}, y^{(i)}\}$ disponíveis: gradiente descendente em **batch**
- $J(\theta)$ na regressão linear poderia ser minimizada analiticamente (veremos mais tarde!)

OBSERVAÇÕES:

- Lembre-se de que a atualização de θ_0 e θ_1 é SIMULTÂNEA
- $J(\theta)$ na regressão linear é **convexa**
- Usa todos os $\{x^{(i)}, y^{(i)}\}$ disponíveis: gradiente descendente em **batch**
- $J(\theta)$ na regressão linear poderia ser minimizada analiticamente (veremos mais tarde!)

OBSERVAÇÕES:

- Lembre-se de que a atualização de θ_0 e θ_1 é **SIMULTÂNEA**
- $J(\theta)$ na regressão linear é **convexa**
- Usa todos os $\{x^{(i)}, y^{(i)}\}$ disponíveis: gradiente descendente em **batch**
- $J(\theta)$ na regressão linear poderia ser minimizada analiticamente (veremos mais tarde!)

OBSERVAÇÕES:

- Lembre-se de que a atualização de θ_0 e θ_1 é SIMULTÂNEA
- $J(\theta)$ na regressão linear é **convexa**
- Usa todos os $\{x^{(i)}, y^{(i)}\}$ disponíveis: gradiente descendente em **batch**
- $J(\theta)$ na regressão linear poderia ser minimizada analiticamente (veremos mais tarde!)

VETORIZAÇÃO:

Defina-se

$$\mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ \vdots & \vdots \\ 1 & x^{(m)} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Então:

$$h_{\theta}(\mathbf{X}) = \mathbf{X}\theta.$$

VETORIZAÇÃO:

Defina-se

$$\mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ \vdots & \vdots \\ 1 & x^{(m)} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Então:

$$h_{\theta}(\mathbf{X}) = \mathbf{X}\theta.$$