

Aula 02

Aprendizado Supervisionado e Regressão Linear

Agenda

01

Aprendizado
Supervisionado

02

Regressão Linear

03

Como utilizar uma
Regressão Linear?

04

Tratando os
dados

05

Implementando uma
regressão linear

06

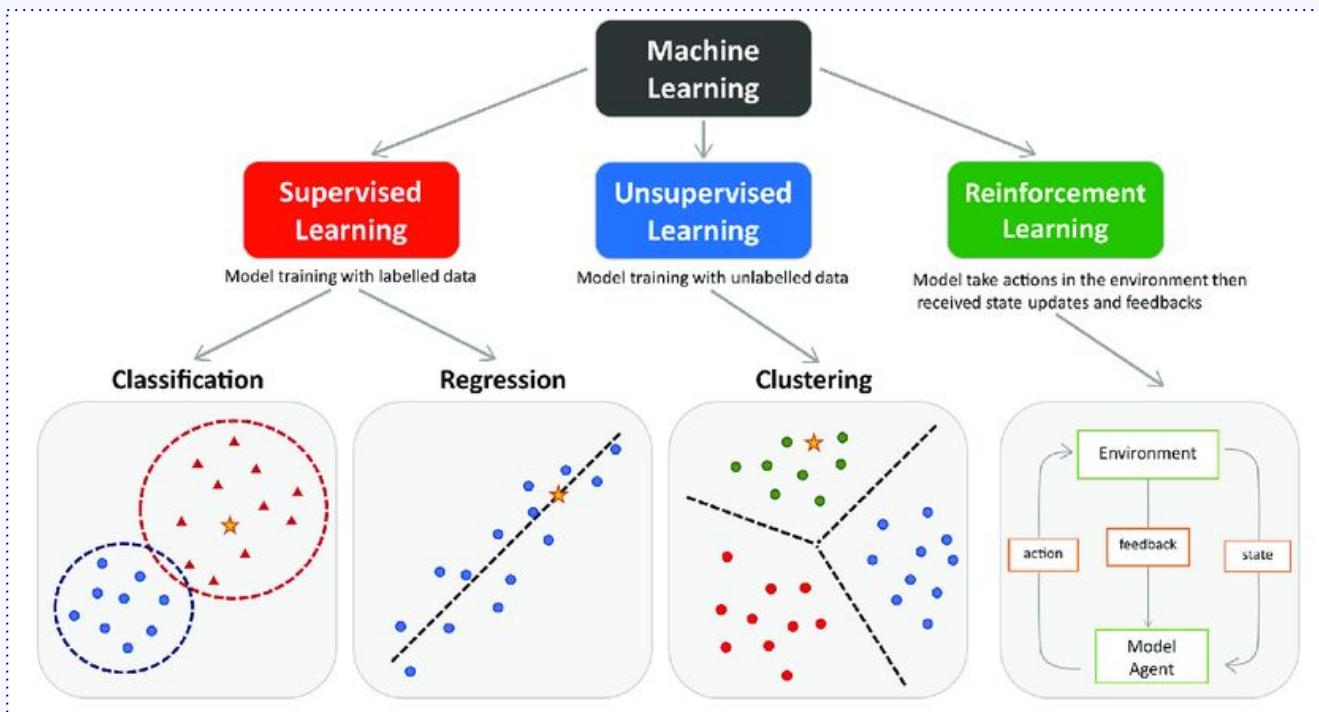
O problema de predição
de preços de casas

01

Aprendizado Supervisionado



Paradigmas de aprendizado



Retirado de:

https://www.researchgate.net/figure/The-main-types-of-machine-learning-Main-approaches-include-classification-and_fig1_354960266

Paradigmas de aprendizado

Aprendizado supervisionado:



O objetivo é “**predizer**” uma saída esperada/alvo.



Saída esperada /alvo é **conhecida**.

Aprendizado Não-supervisionado:



O objetivo é **encontrar padrões** na estrutura dos dados.



Não há uma saída esperada /alvo **conhecida**.

Aprendizado Auto-supervisionado:



O objetivo é extrair e **capturar os padrões** nos dados.



A saída esperada /alvo **faz parte dos dados de entrada**.

Como é o aprendizado supervisionado?

Dados sem
a resposta

Processamento
do modelo

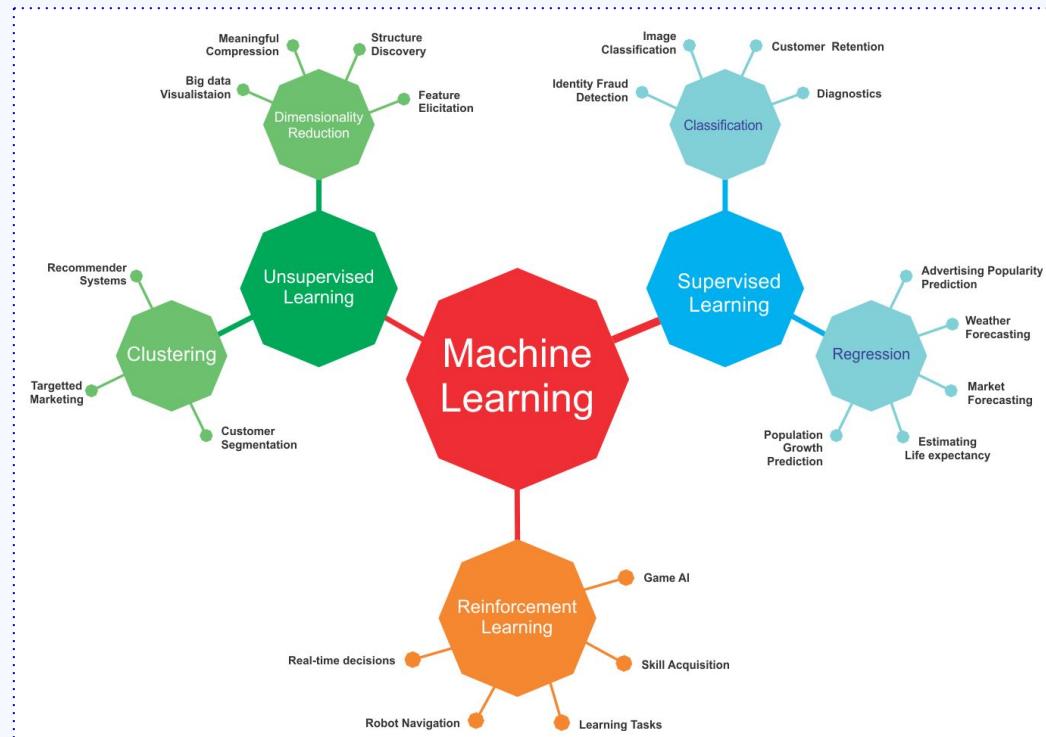
Chute de
resposta

Ajuste das
regras

O chute
está certo
ou errado?



Tarefas (ou “problemas”) em ML



Retirado de: <https://subscription.packtpub.com/book/data/9781789345070/1/ch01lvl1sec04/ml-tasks>

Tarefas supervisionadas

Regressão

O objetivo é **quantificar e inferir a relação** de uma **variável dependente** (*variável de resposta*) com **variáveis independentes** (*variáveis explicativas*).

Classificação

Identificar **a que categoria pertence uma nova observação, com base em um conjunto de dados** contendo observações cujas suas categorias são conhecidas.

Tipos de regressão

- Regressão Linear
- Regressão Não-linear
- Regressão Polinomial

Regressão Logística



02

Regressão Linear



Motivação



\$150.000



\$385.000

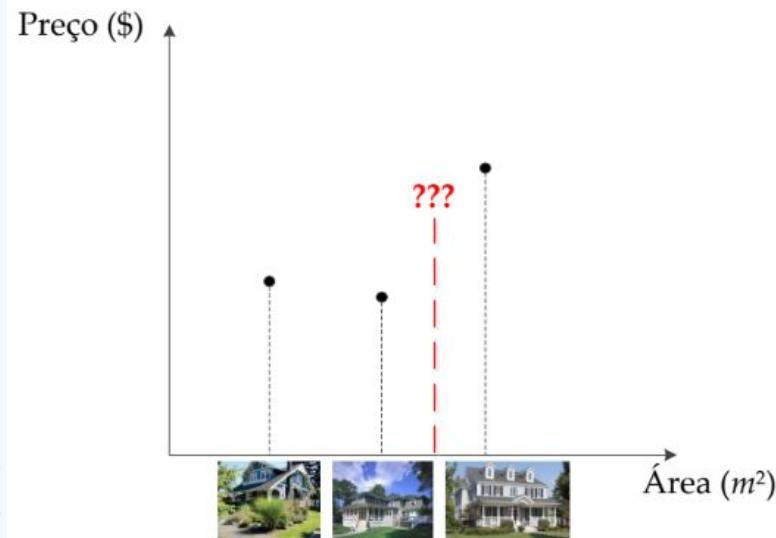


????

Hipótese: Qual seria a relação matemática entre o valor do imóvel e a sua área?

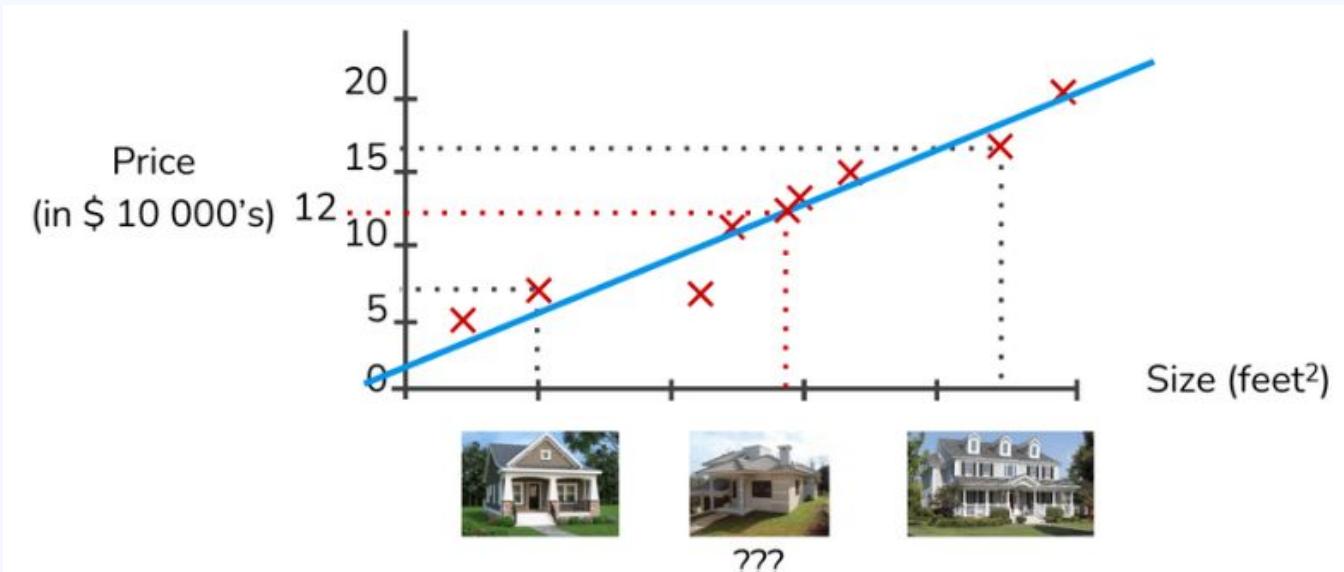
Motivação

Desafio: Ao saber a área e valor de diferentes imóveis, o objetivo é usar a relação matemática entre as grandezas envolvidas para, então, estimar o valor de novos imóveis



Definição de problema

Problema: Consiste em achar o valor ótimo com base nos parâmetros do modelo, ou seja, qual a melhor estimativa de preço de uma casa.



O que é uma Regressão Linear?

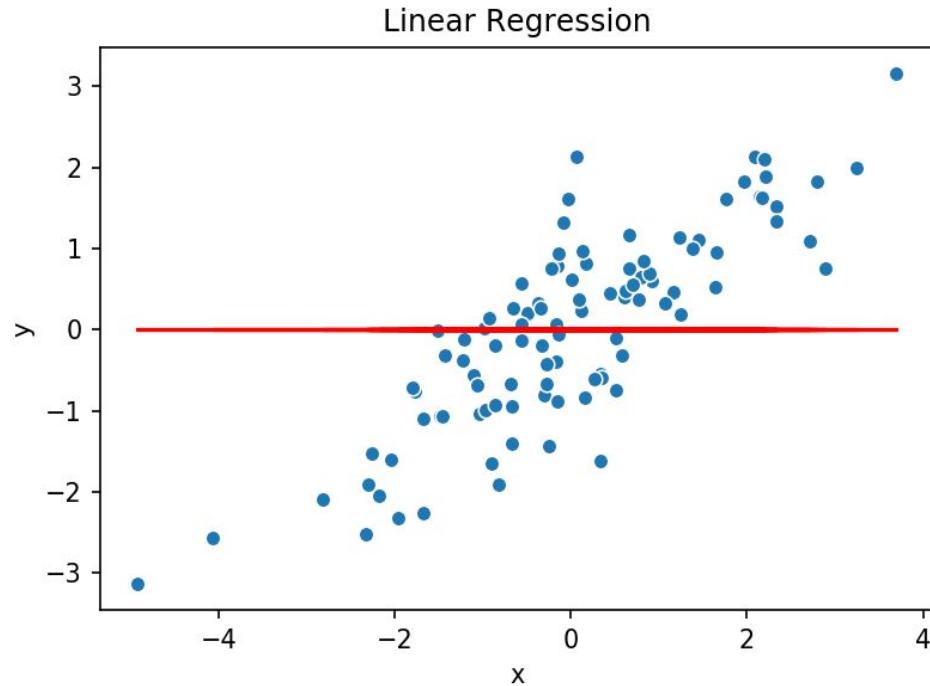
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Uma função F é linear se:

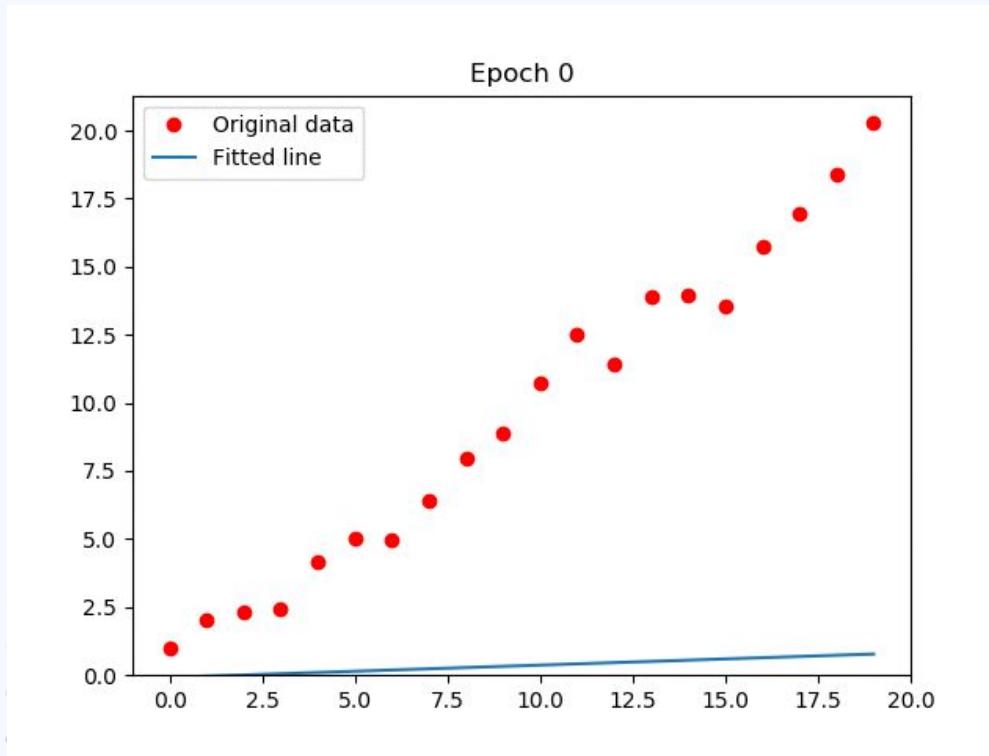
$$F(x + y) = F(x) + F(y) \quad \text{e} \quad c * F(x) = F(c * x)$$

Pode ser representada por uma reta, um plano, ...

O que é uma Regressão Linear?



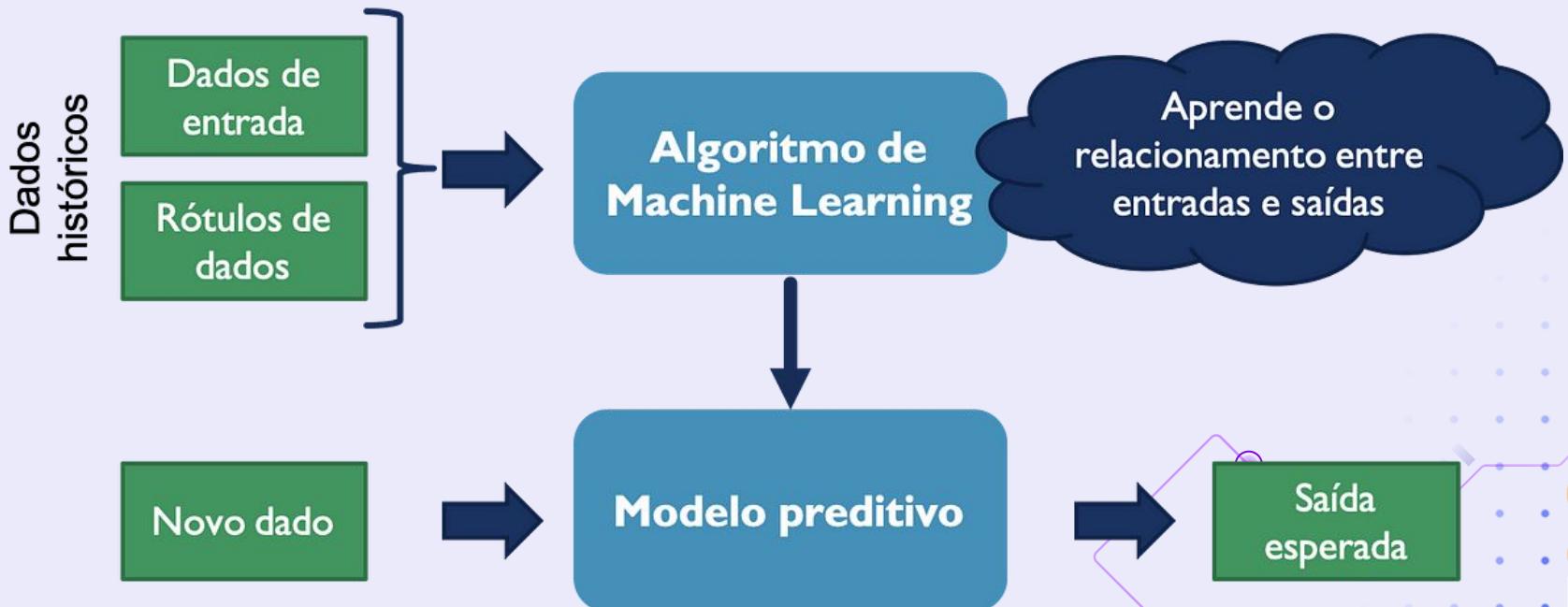
O que é uma Regressão Linear?



O que é um modelo?



Modelando um fenômeno



Quão bom é o nosso modelo?

Como esse ajuste é calculado?

Podemos metrificar o quão bom o nosso modelo é?

Qual é o **erro** dele?

O erro do modelo é uma espécie de diferença entre o chute do modelo e a resposta correta.

Quão bom é o nosso modelo?

Como esse ajuste é calculado?

Vamos pensar no nosso erro como **um custo** ou **uma perda**.

Queremos minimizar o custo!

Objetivo: minimizar uma função que representa o custo do nosso modelo.

Qual a Função de Custo?

A nossa métrica de erro será a *Mean Squared Error*:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Função de Custo

Hypothesis:

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

Parameters:

$$\theta_0, \theta_1$$



Cost Function:

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

Goal:

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

Simplified

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



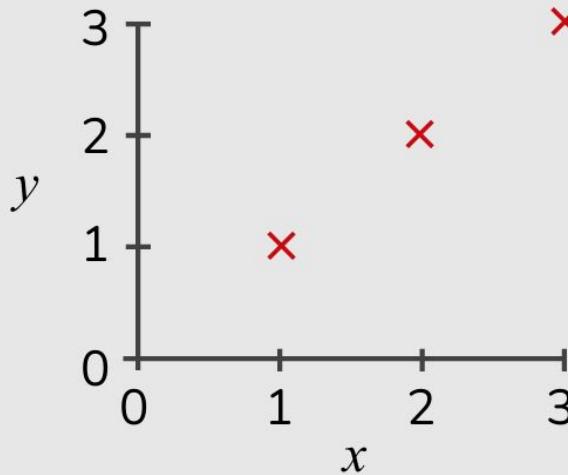
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

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

Função de Custo

$$h_{\theta}(x)$$

(for fixed θ_1 , this is a function of x)



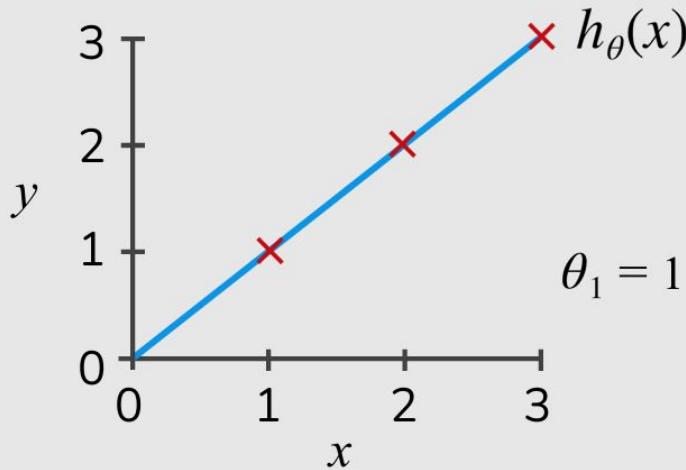
$$J(\theta_1)$$

(function of the parameters θ_1)

Função de Custo

$$h_{\theta}(x)$$

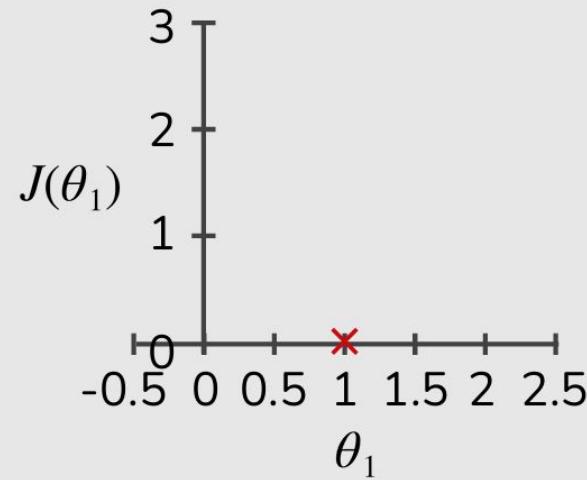
(for fixed θ_1 , this is a function of x)



$$J(\theta_1) = J(1) = 0$$

$$J(\theta_1)$$

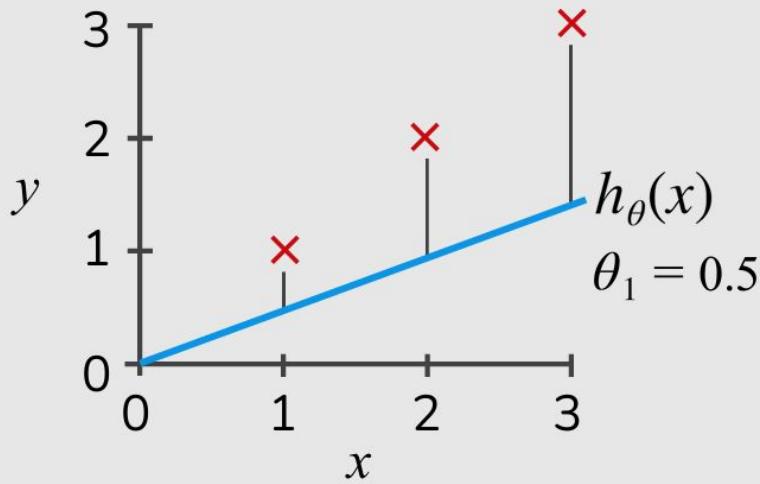
(function of the parameters θ_1)



Função de Custo

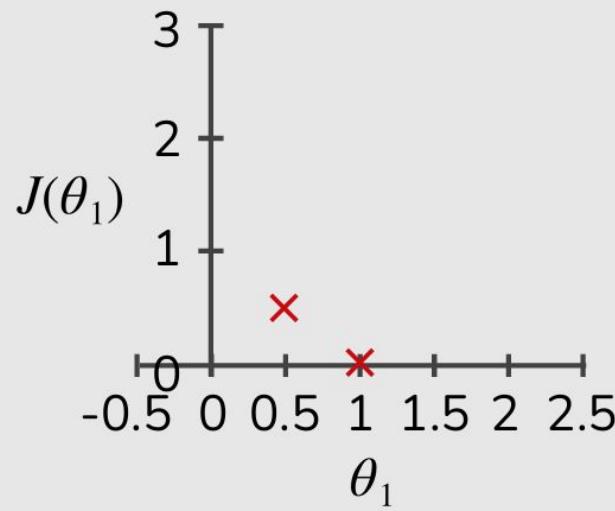
$$h_{\theta}(x)$$

(for fixed θ_1 , this is a function of x)



$$J(\theta_1)$$

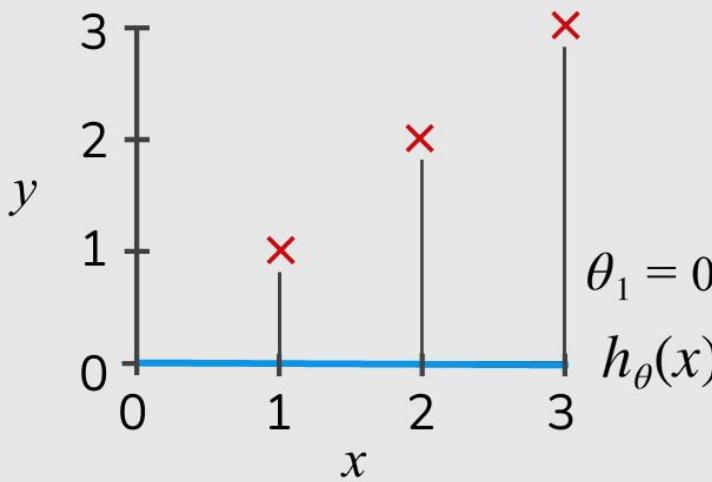
(function of the parameters θ_1)



Função de custo

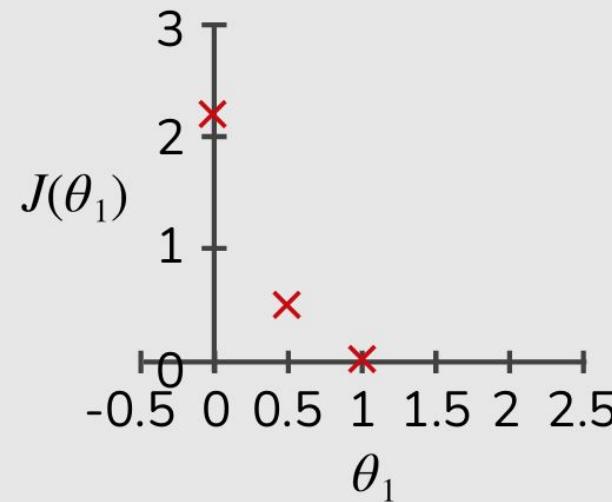
$$h_{\theta}(x)$$

(for fixed θ_1 , this is a function of x)



$$J(\theta_1)$$

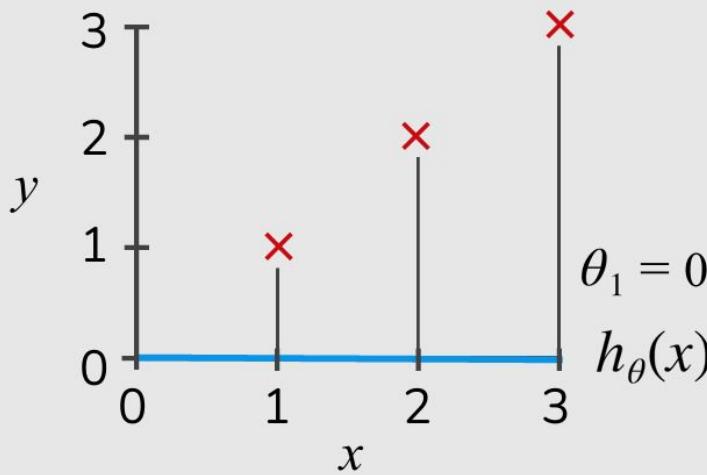
(function of the parameters θ_1)



Função de Custo

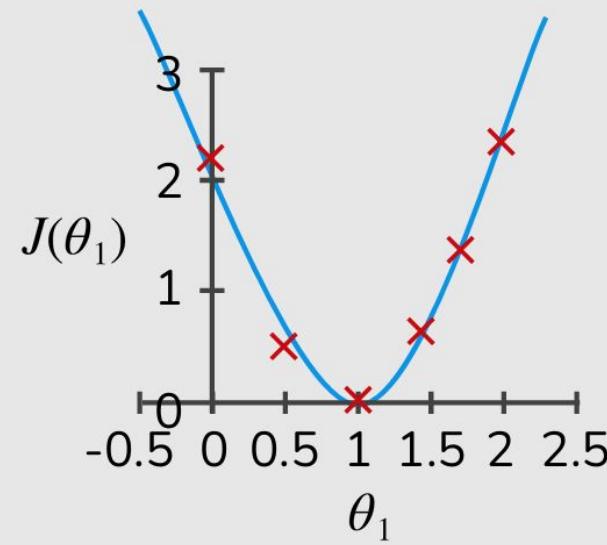
$$h_\theta(x)$$

(for fixed θ_1 , this is a function of x)



$$J(\theta_1)$$

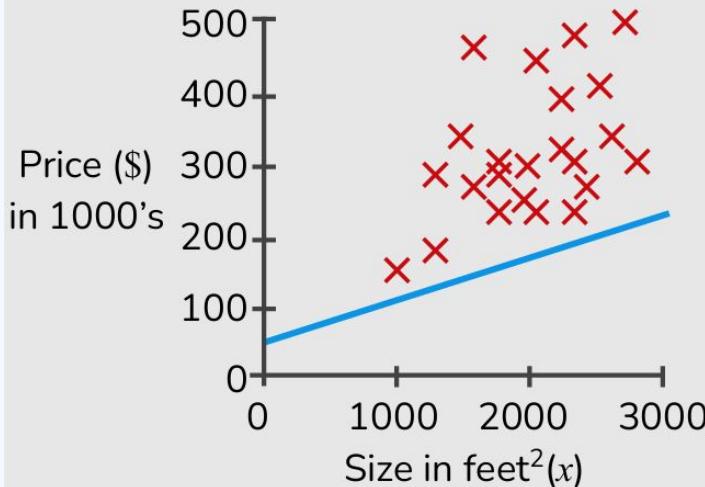
(function of the parameters θ_1)



Função de Custo

$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)

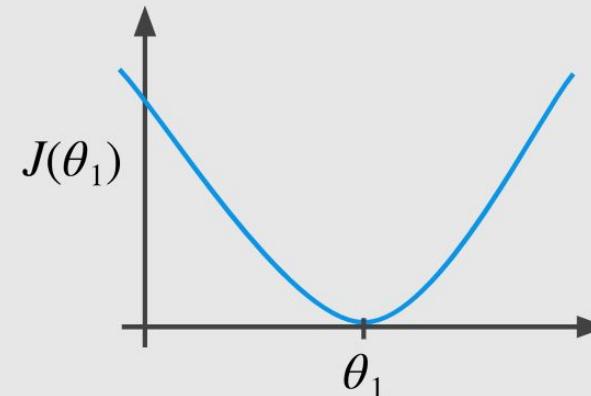


$$h_{\theta}(x) = 50 + 0.06x$$

$$\begin{aligned}\theta_0 &= 50 \\ \theta_1 &= 0.06\end{aligned}$$

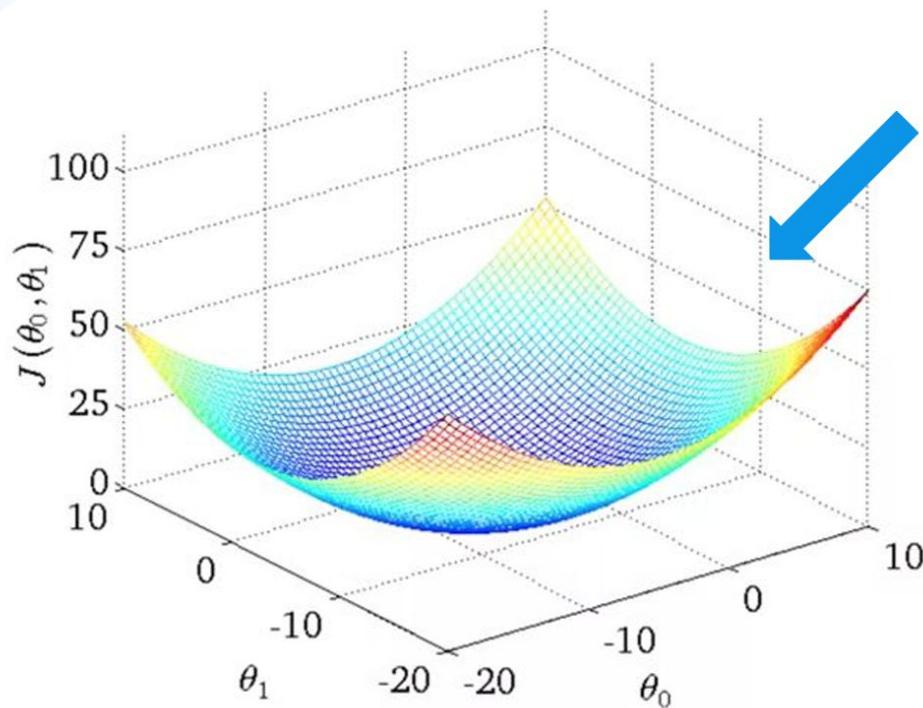
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



θ_0 and θ_1 ?

Função de Custo

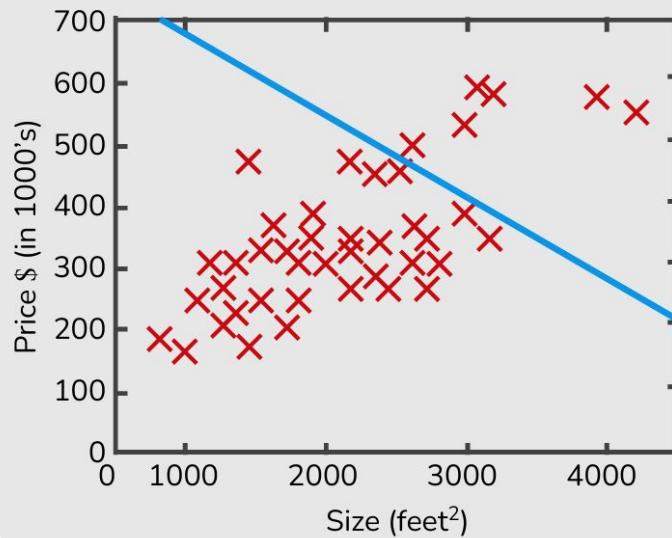


Convex
Function

Função de Custo

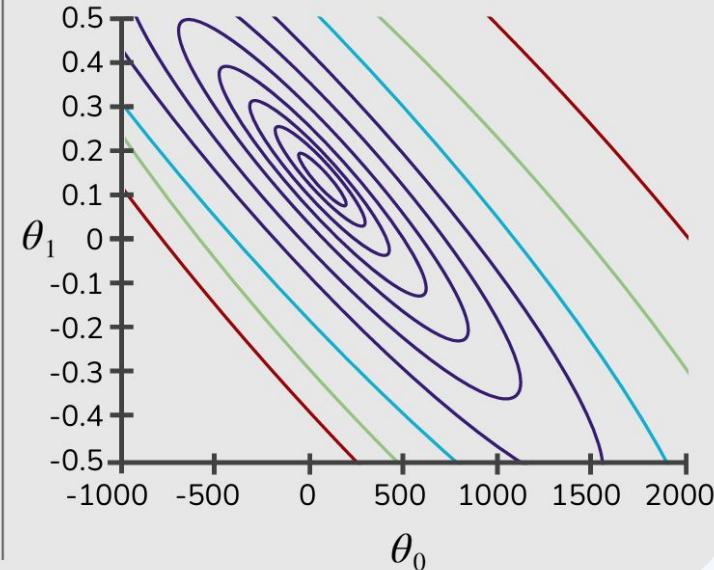
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



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

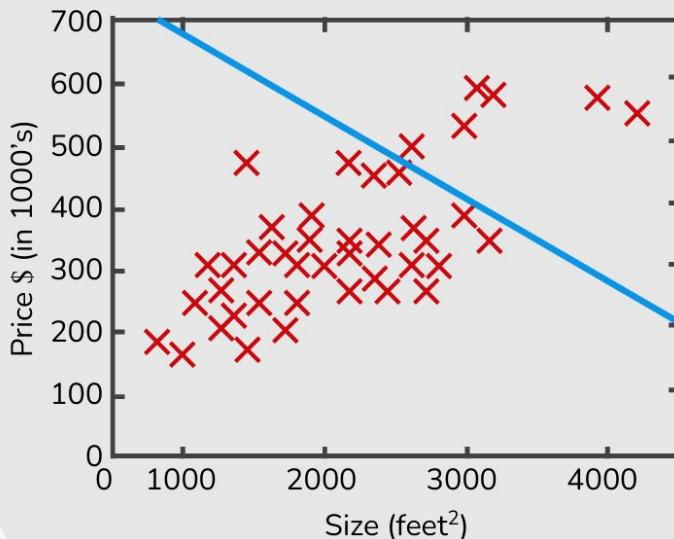
(function of the parameters θ_0, θ_1)



Função de Custo

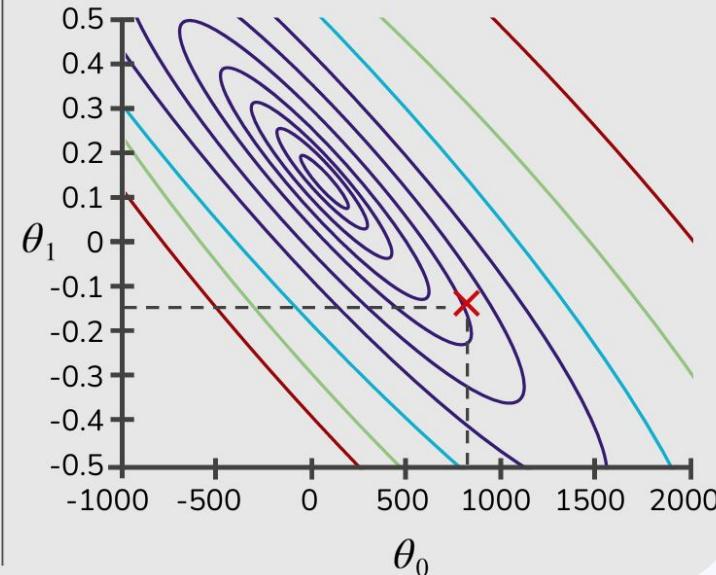
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



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

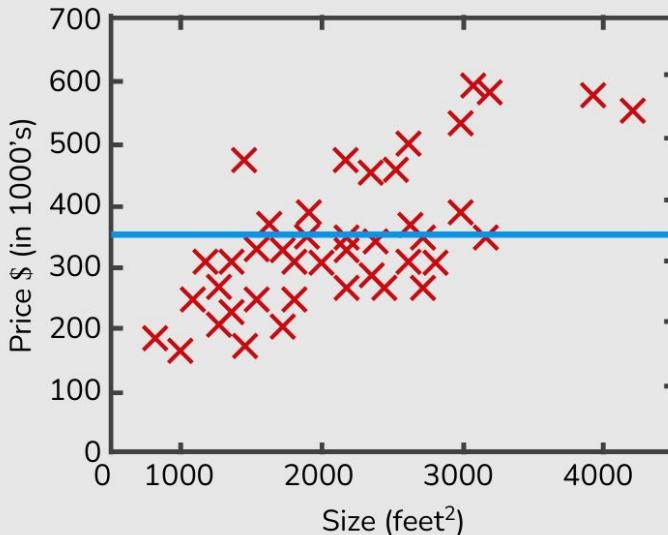
(function of the parameters θ_0, θ_1)



Função de Custo

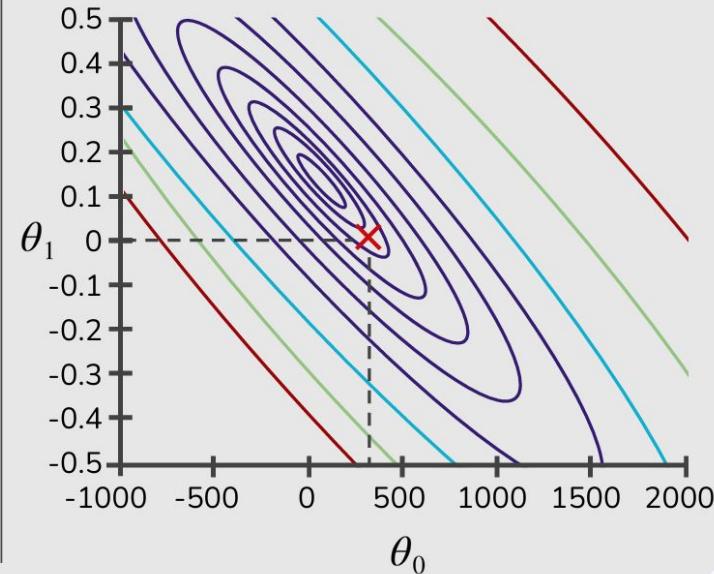
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



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

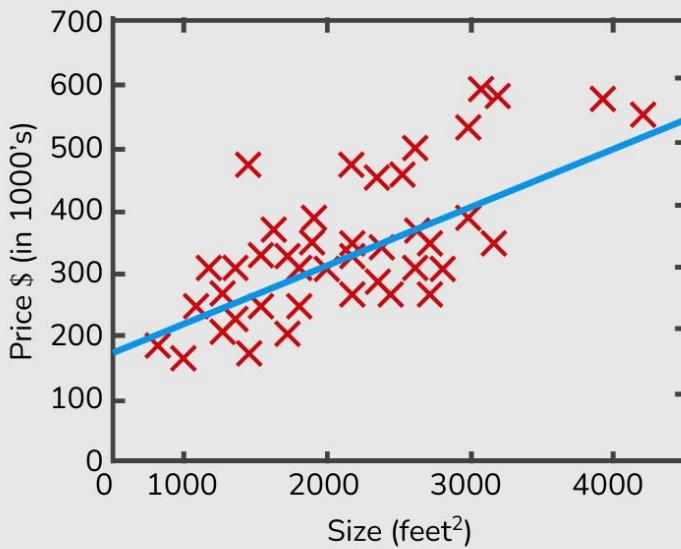
(function of the parameters θ_0, θ_1)



Função de Custo

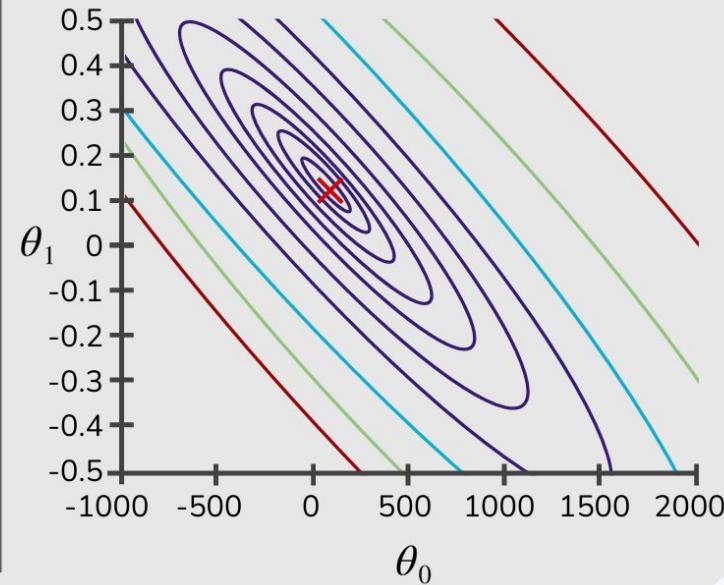
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



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

(function of the parameters θ_0, θ_1)



A Visual Introduction To (Almost) Everything You Should Know



LINEAR REGRESSION

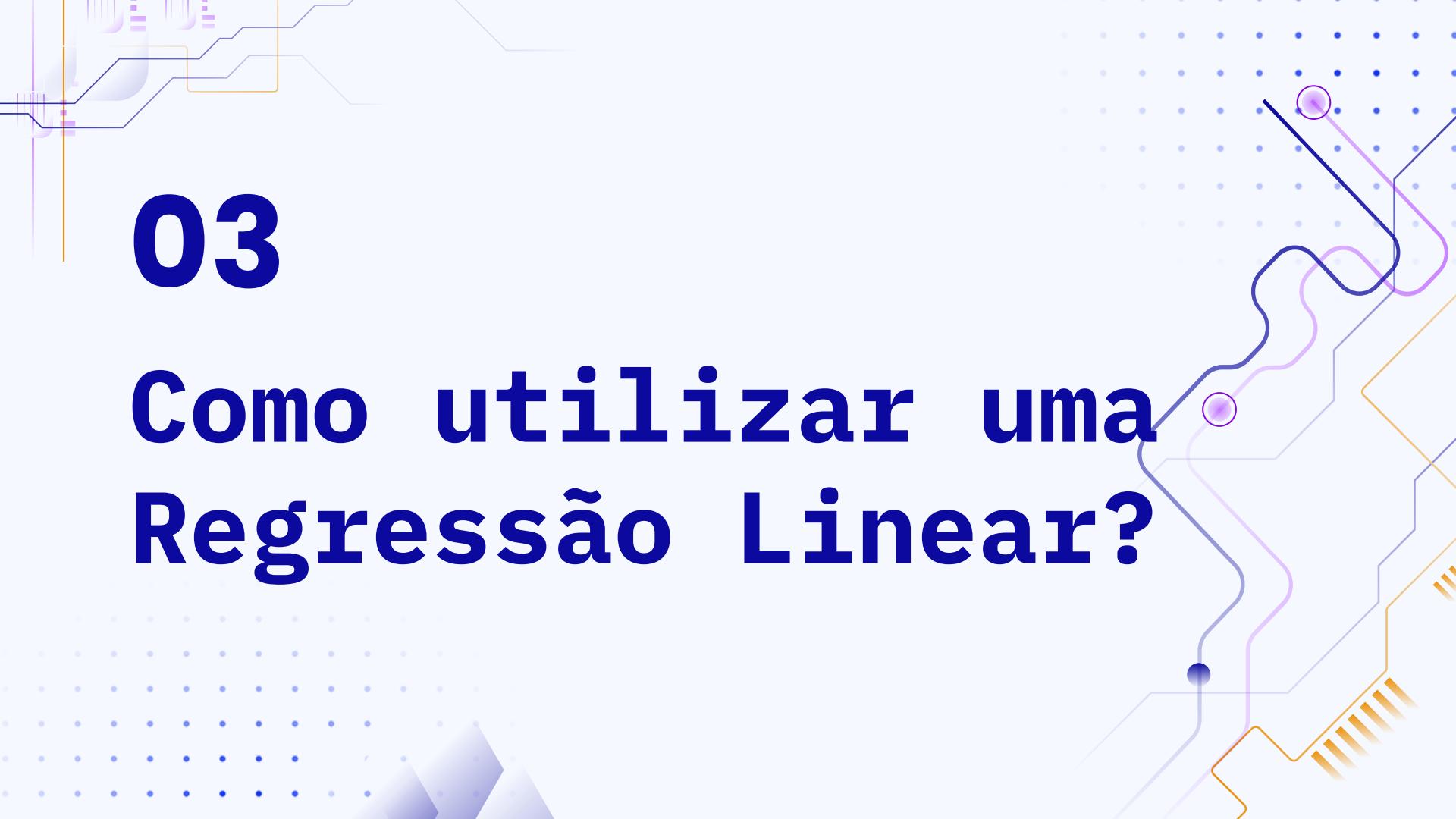
A Visual Introduction To (Almost)
Everything You Should Know

Jared Wilber, September 2022

Linear Regression is a simple and powerful model for predicting a numeric response from a set of one or more independent variables. This article will focus mostly on how the method is used in machine learning, so we won't cover common use cases like causal inference or experimental design. And although it may seem like linear regression is overlooked in modern machine learning's ever increasing world of complexity, it remains a valuable tool for understanding data and making predictions.

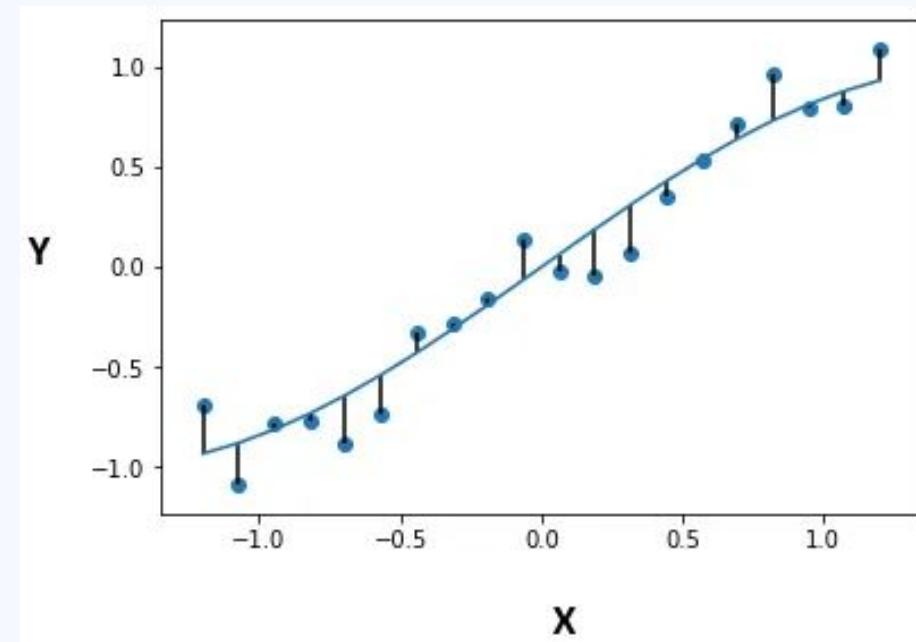
03

Como utilizar uma Regressão Linear?

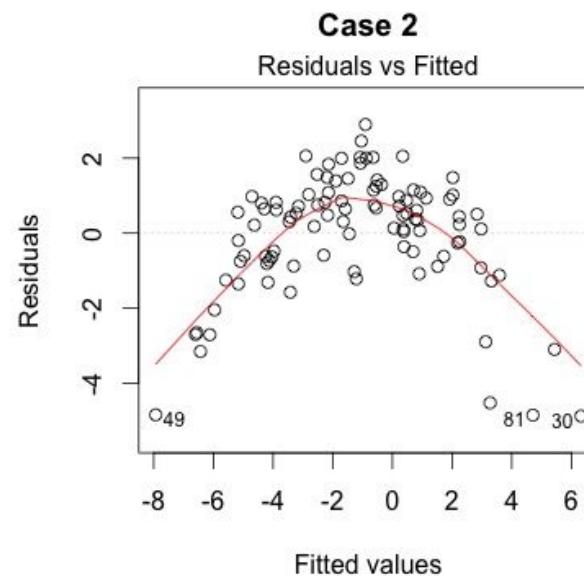
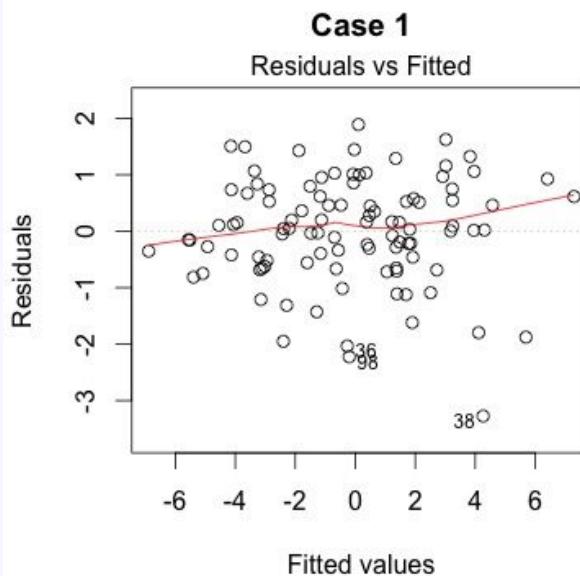


Como saber se seus dados tem uma relação linear?

Checar o erro ou o resíduo em função das previsões do seu modelo!



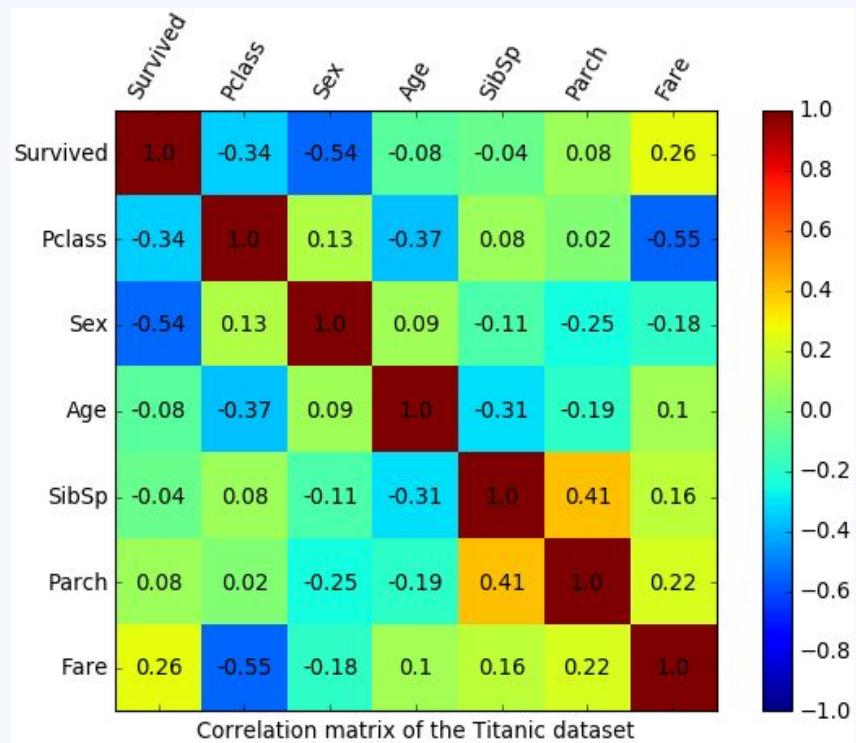
Como saber se seus dados tem uma relação linear?



Quais *features* escolher?

Correlação de Pearson.

Correlação de Spearman.



04

Tratando os dados



Variáveis categóricas

Como lidar com variáveis categóricas, ou seja, que possuem valores discretos?

Ex: *feature* "Animal"

Salário	Altura	Animal	Comprou
2630,00	170	Gato	0
5650,00	169	Cachorro	1
1250,00	185	Peixe	1
8900,00	176	Cachorro	0

Variáveis categóricas

One-hot encoding:

Salário	Altura	Animal Gato	Animal Cachorro	Animal Peixe	Comprou
2630,00	170	1	0	0	0
5650,00	169	0	1	0	1
1250,00	185	0	0	1	1
8900,00	176	0	1	0	0

Variáveis categóricas

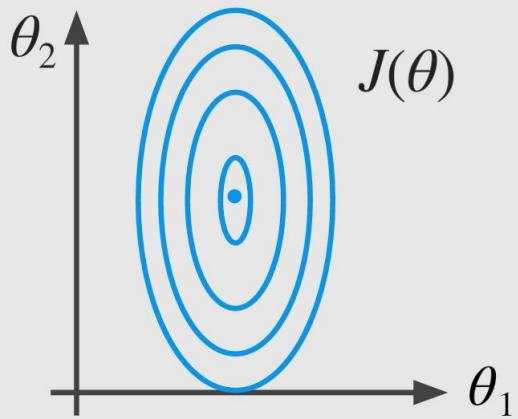
Label encoding:

Salário	Altura	Animal	Comprou
2630,00	170	0	0
5650,00	169	1	1
1250,00	185	2	1
8900,00	176	1	0

Normalizando as *features*

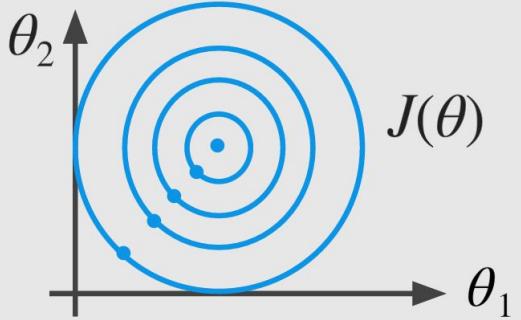
x_1 = size (0–2000 feet²)

x_2 = number of bedrooms (1–5)



$$x_1 = \frac{\text{size (feet}^2)}{2000}$$

$$x_2 = \frac{\text{number of bedrooms}}{5}$$

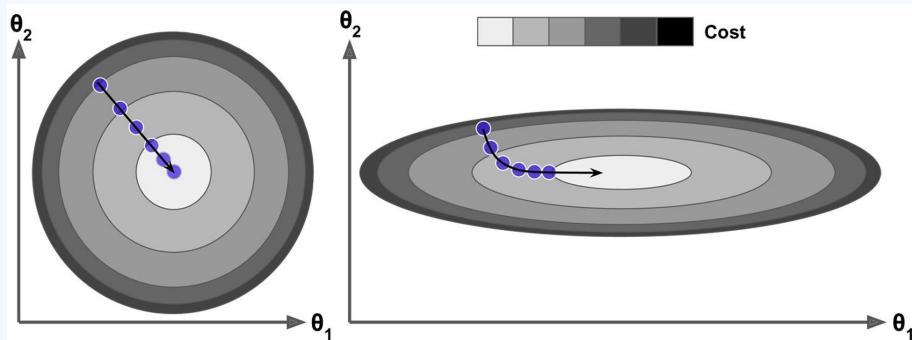


Normalizando as *features*

Para lidar com *features* que preencham intervalos muito diferentes, podemos usar:

Min-Max Scaling

$$x_j^{(i)} \leftarrow \frac{x_j^{(i)} - \min(x_j)}{\max(x_j) - \min(x_j)}$$



05

Implementando uma regressão linear

+ Create



ASHISH · 5Y AGO · 174,629 VIEWS

▲ 258

Copy & Edit 3393



Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

VIEWED

View Active Events

Sales Prediction (Simple Linear Regression)

Python · [Advertising Dataset](#)

Notebook Input Output Logs Comments (14)

Run

14.2s

Version 1 of 1

Beginner

Linear Regression

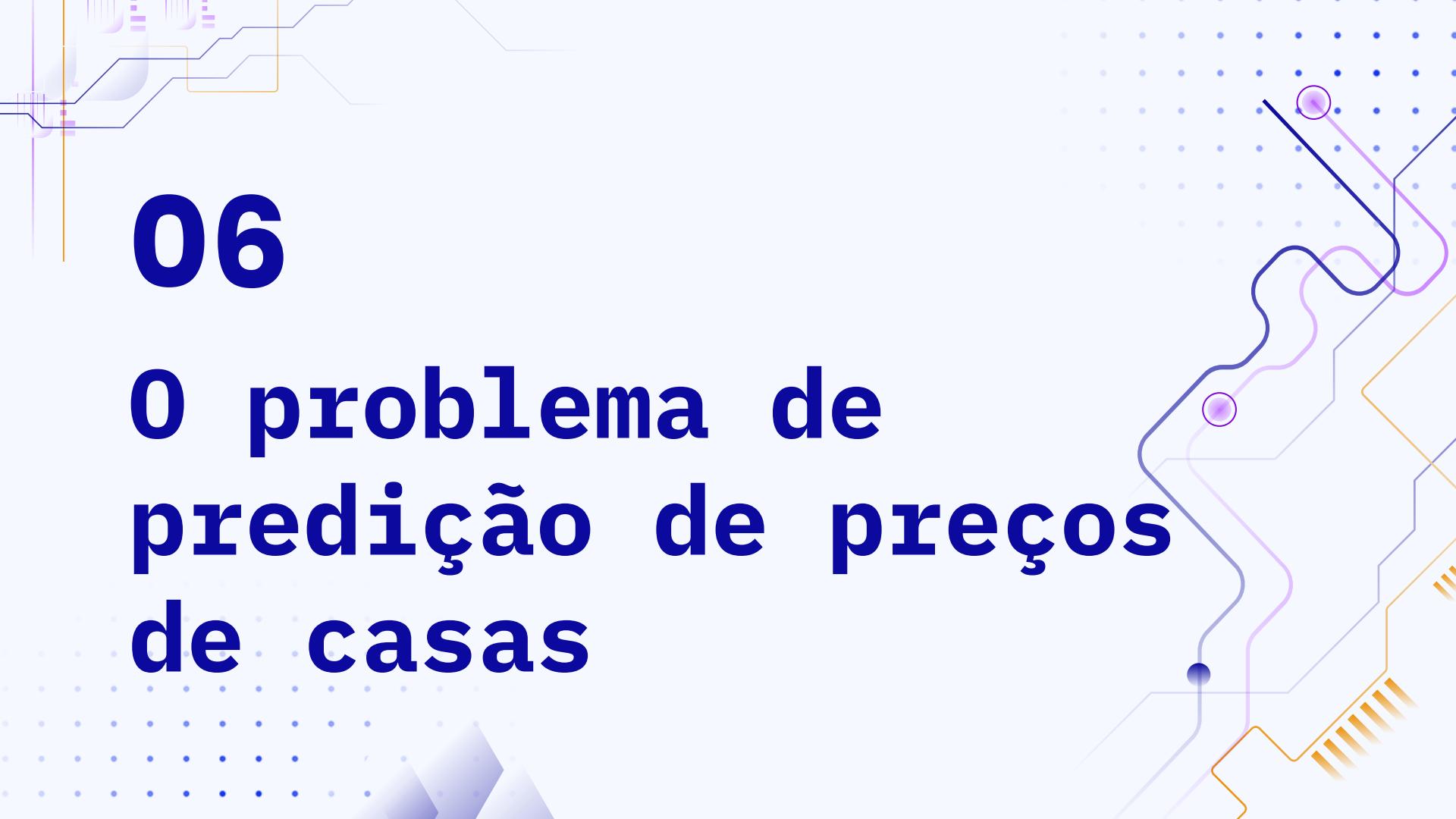


Sales Prediction

(Simple Linear Regression)

06

O problema de predição de preços de casas





Obrigado pessoal!

Até próxima aula :)



Iris Data Science UNICAMP



@irisdatascienceunicamp

Referências

Parte do material foi adaptado dos slides da Prof^a. Sandra Avila apresentados na disciplina MC886.

Material online:

- Linear Regression: A Visual Introduction To (Almost) Everything You Should Know (página web)
- Introdução à Data Science com Python - Data ICMC (repositório no GitHub)

Livros utilizados:

- An Introduction to Statistical Learning (2023)
- Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow (2017)