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

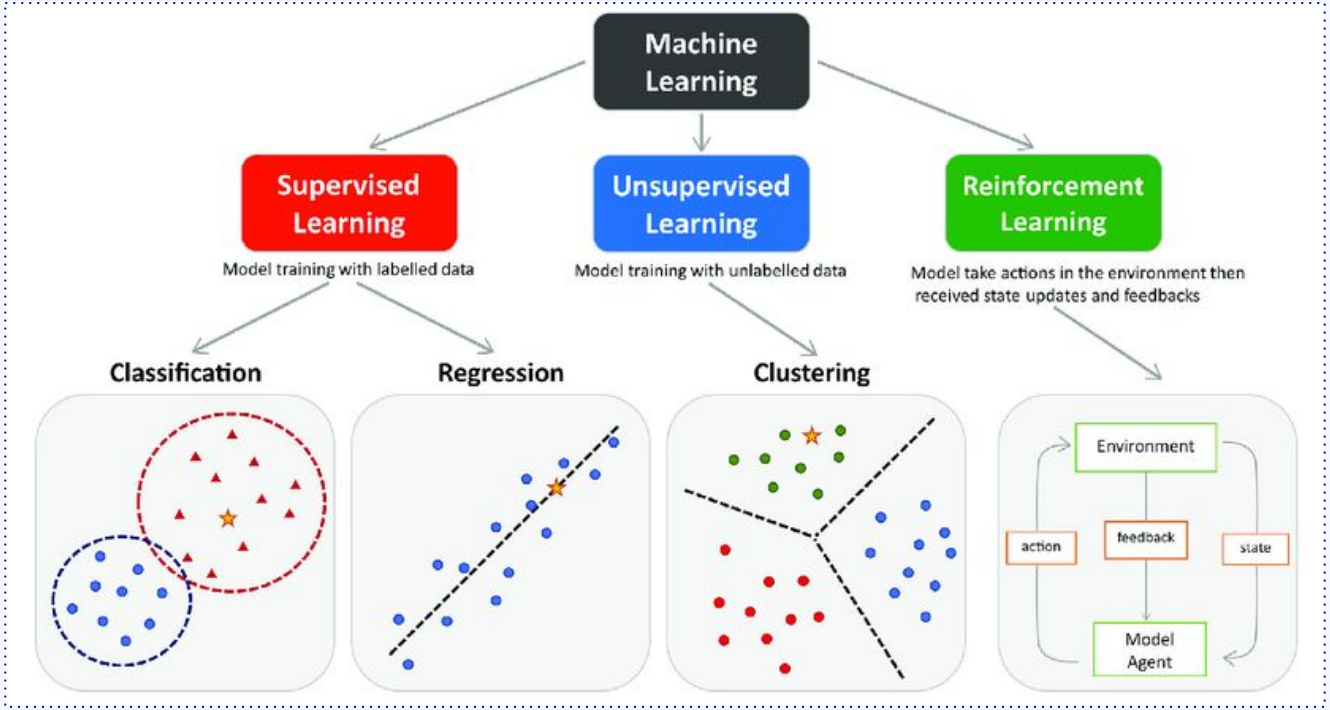
06

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?

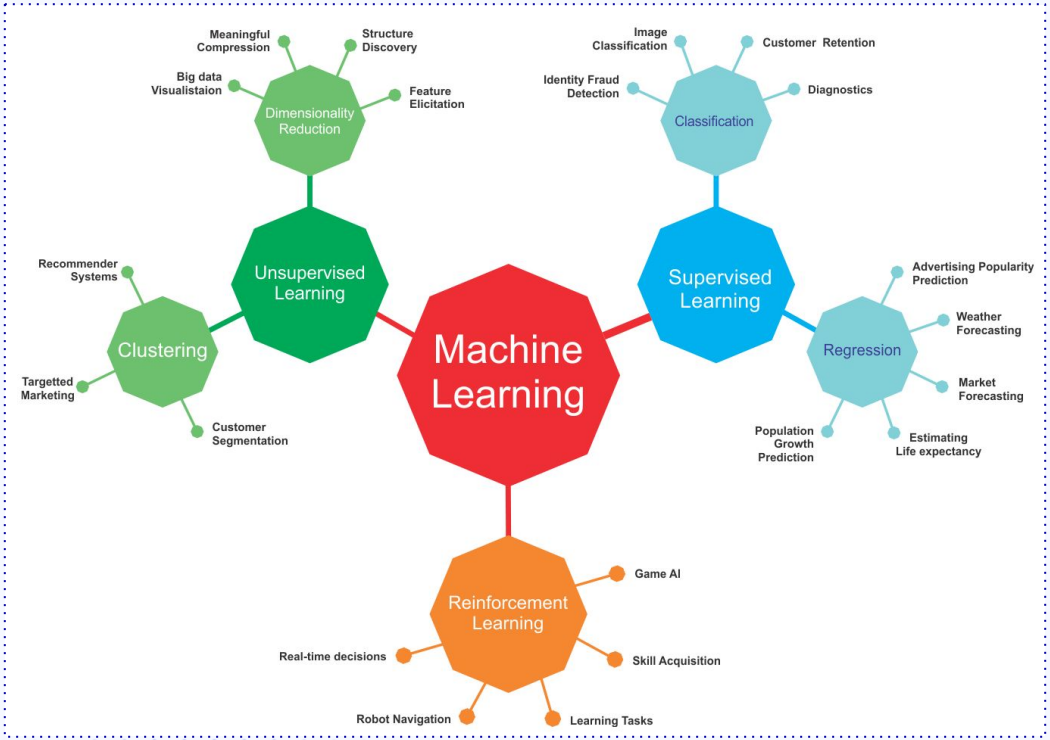
Dados sem a resposta → Processamento do modelo → Chute de resposta

Chute de resposta → O chute está certo ou errado?

O chute está certo ou errado? → Ajuste das regras

Ajuste das regras → Processamento do modelo

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 Polinomial**
- **Regressão Não-linear**
- **Regressão Logística**

02

Regressão Linear

Motivação

$150.000

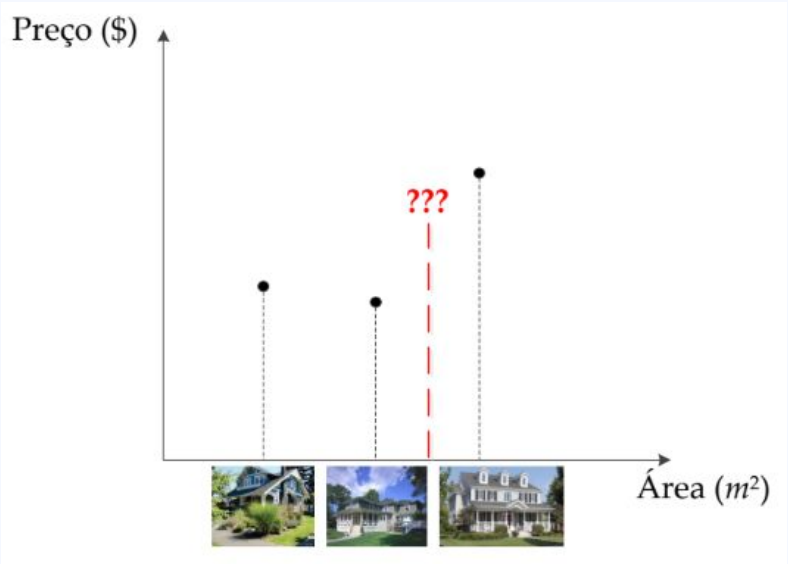
$385.000

????

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

Motivação

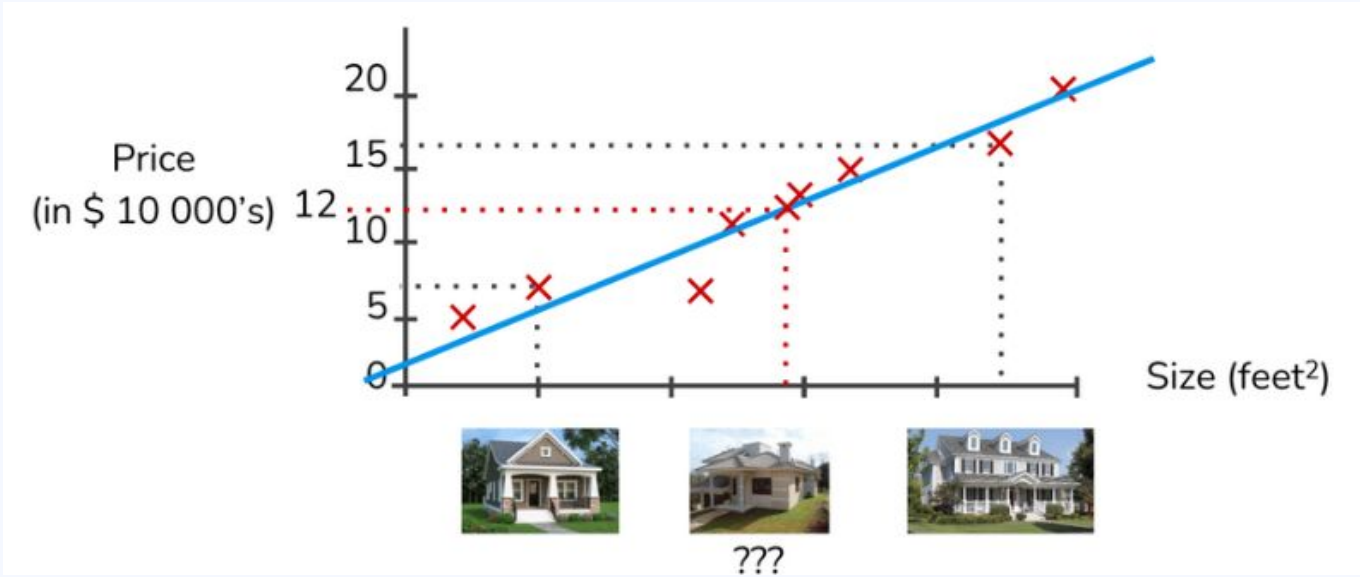
Desafio: Ao saber a área e valor de diferentes imóveis, o objetivo é usar a relação matemática entre as grandezas envolvidas para, então, estimar o valor de novos imóveis



https://www.kaggle.com/datasets/harlfoxem/housesalesprediction

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.



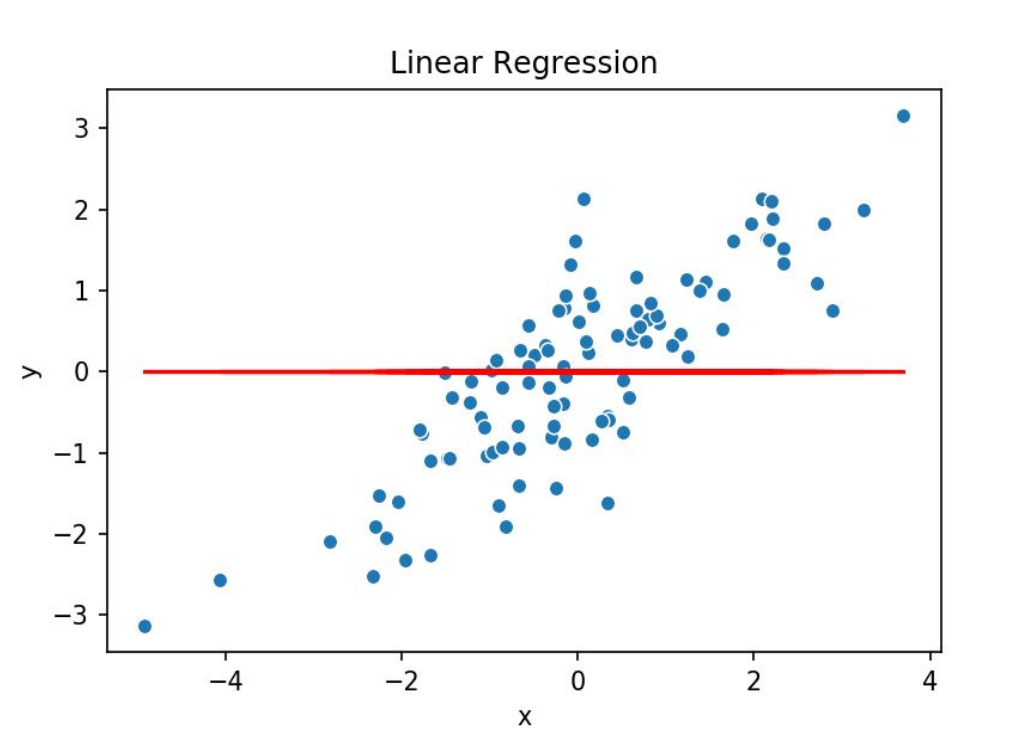
O que é uma Regressão Linear?

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

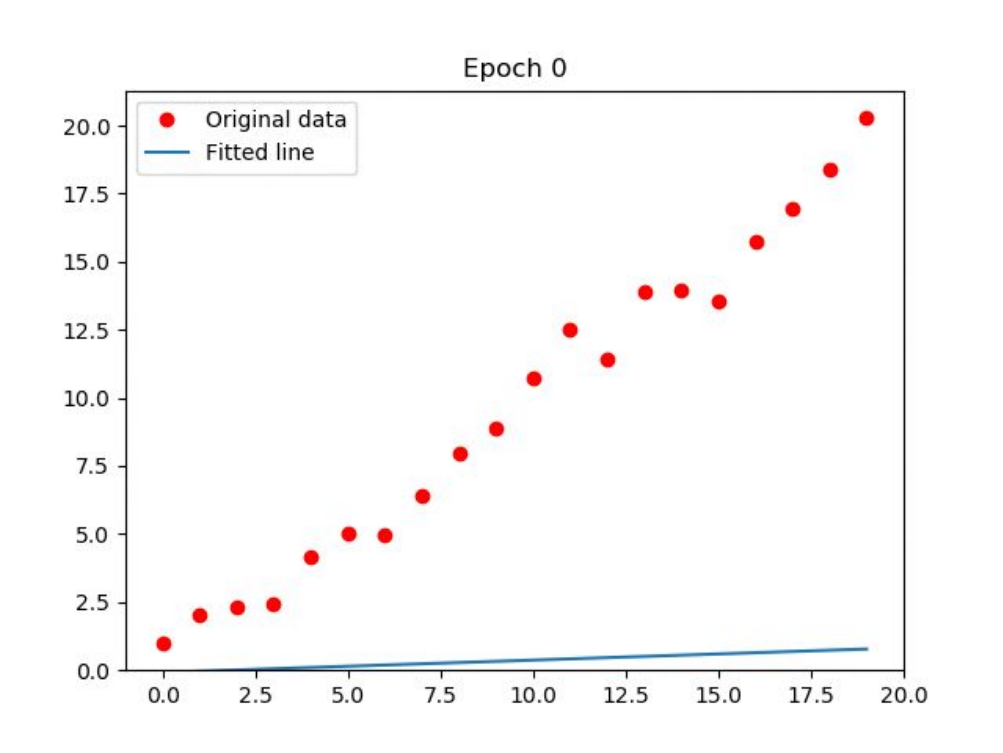
Uma função F é linear se:
*F(x + y) = F(x) + F(y) e c*F(x) = F(c*x)*

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

O que é uma Regressão Linear?

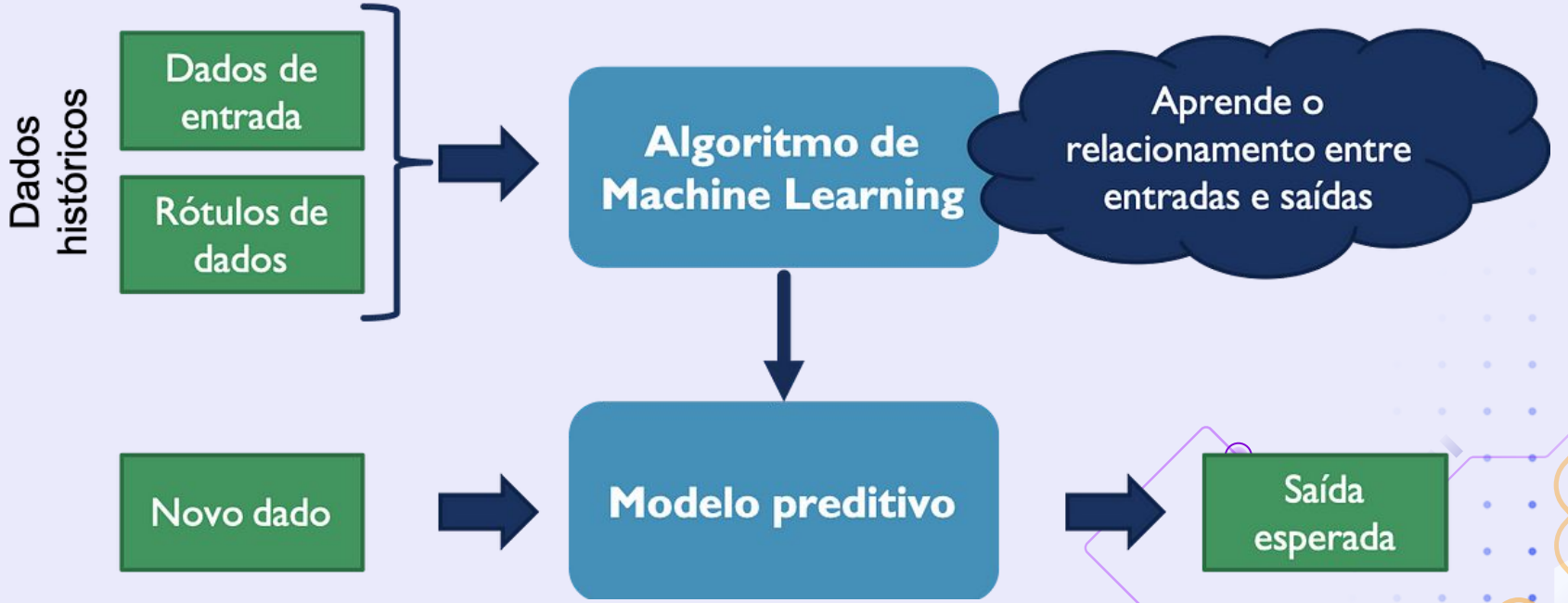


O que é uma Regressão Linear?



O que é um modelo?

Modelando um fenômeno



Retirado de: https://tatianaesc.medium.com/machine-learning-conceitos-e-modelos-f0373bf4f445

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*:

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

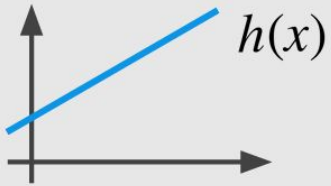
Função de Custo

Hypothesis:

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

Parameters:

$$\theta_0, \theta_1$$



Cost Function:

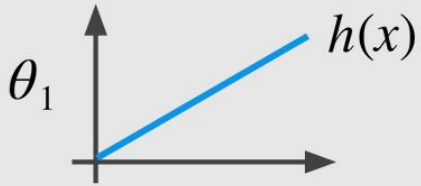
$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Goal:

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

Simplified

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



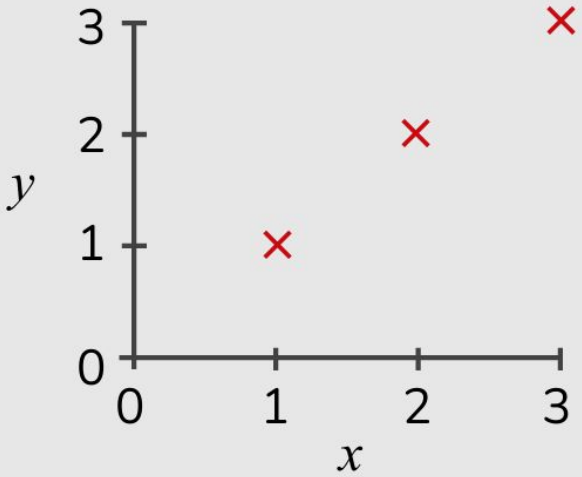
$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

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

Função de Custo

$h_\theta(x)$

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



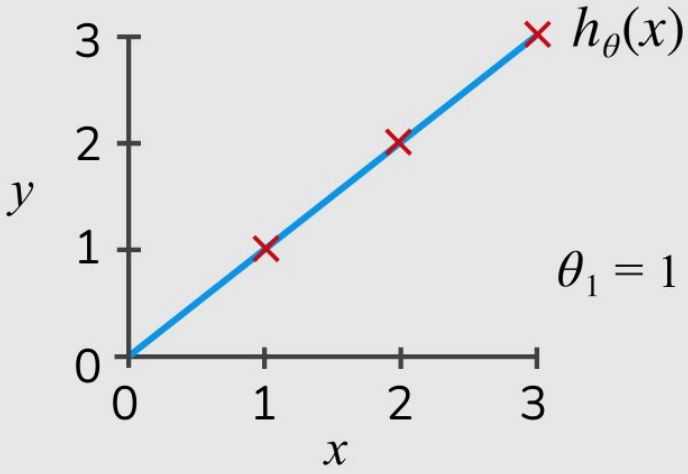
$J(\theta_1)$

(function of the parameters θ_1)

Função de Custo

$h_\theta(x)$

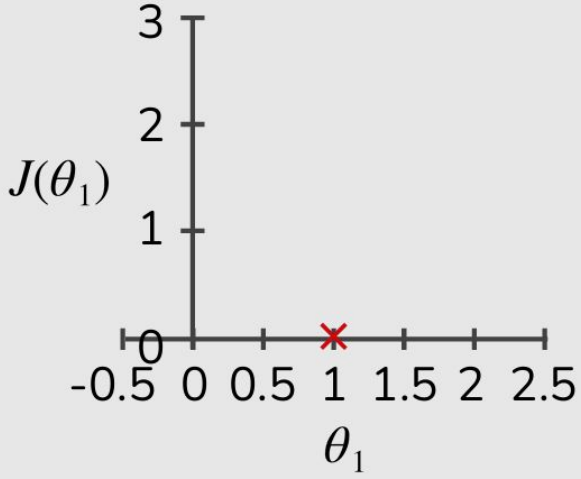
(for fixed θ_1, this is a function of x)



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

$J(\theta_1)$

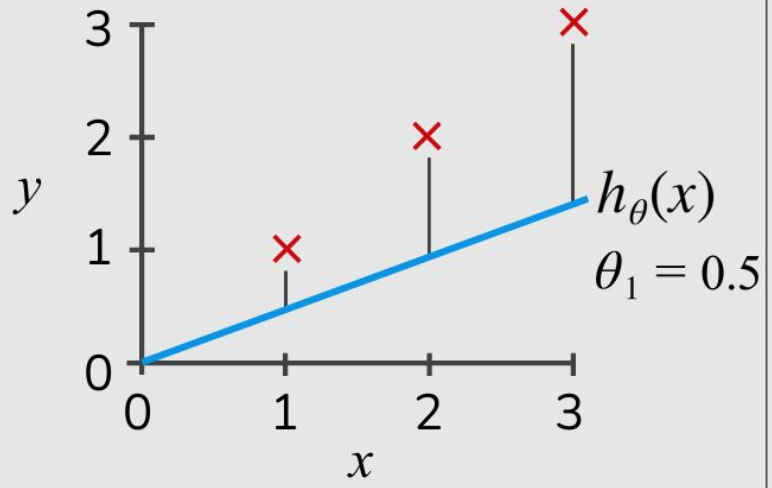
(function of the parameters θ_1)



Função de Custo

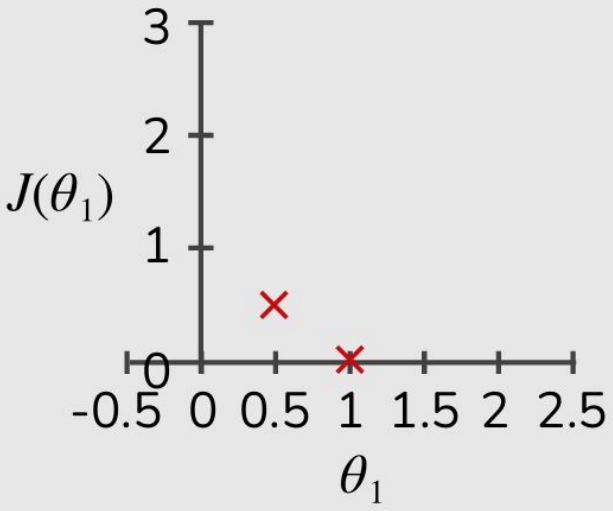
$h_\theta(x)$

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



$J(\theta_1)$

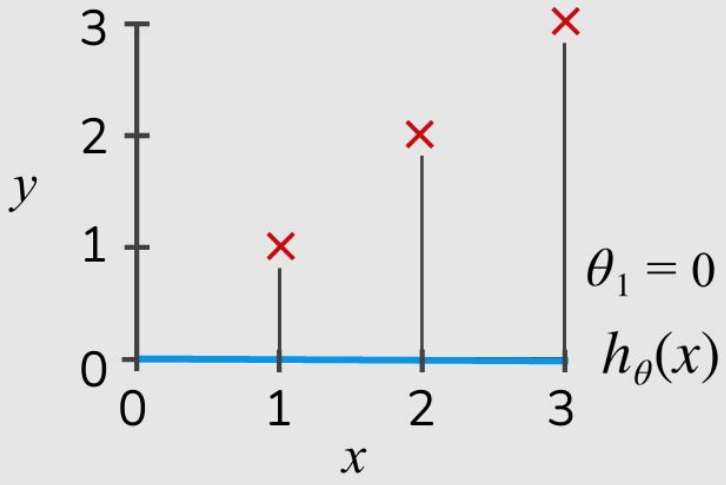
(function of the parameters θ_1)



Função de custo

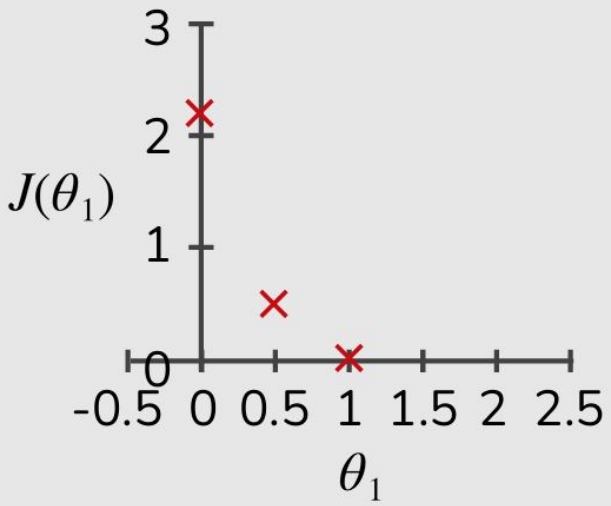
$h_\theta(x)$

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



$J(\theta_1)$

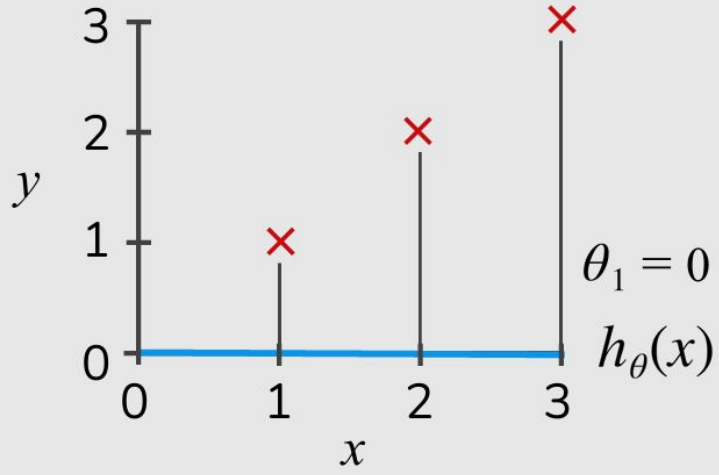
(function of the parameters θ_1)



Função de Custo

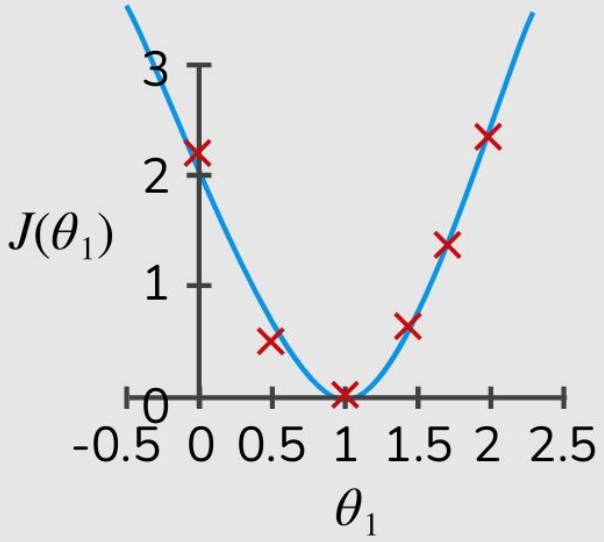
$h_\theta(x)$

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



$J(\theta_1)$

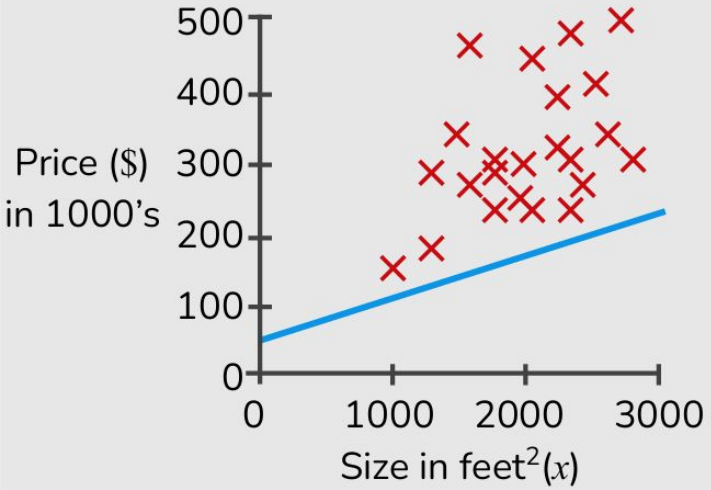
(function of the parameters θ_1)



Função de Custo

$h_\theta(x)$

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



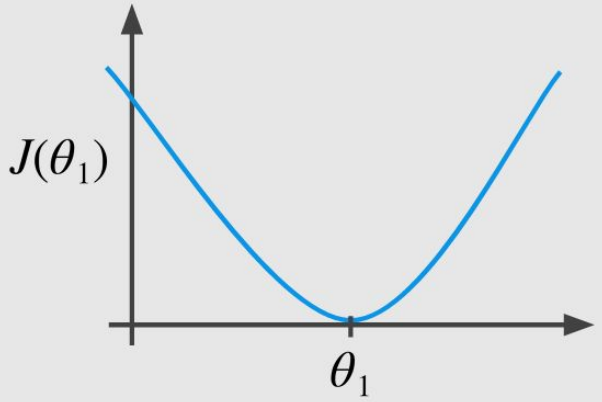
$h_\theta(x) = 50 + 0.06x$

$\theta_0 = 50$

$\theta_1 = 0.06$

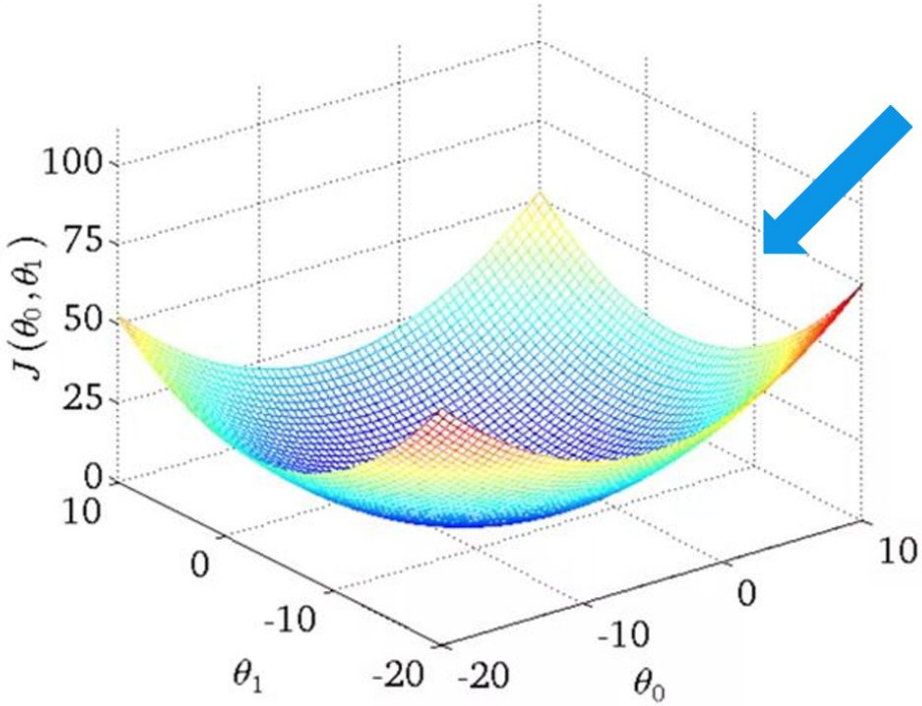
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



θ_0 and θ_1?

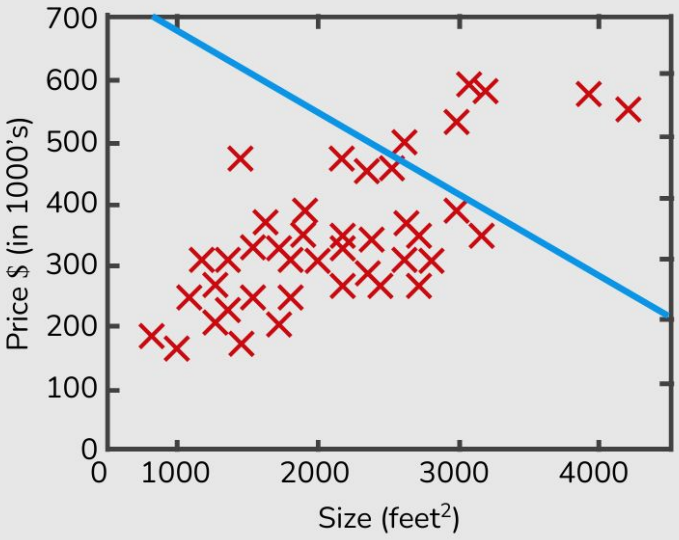
Função de Custo



Função de Custo

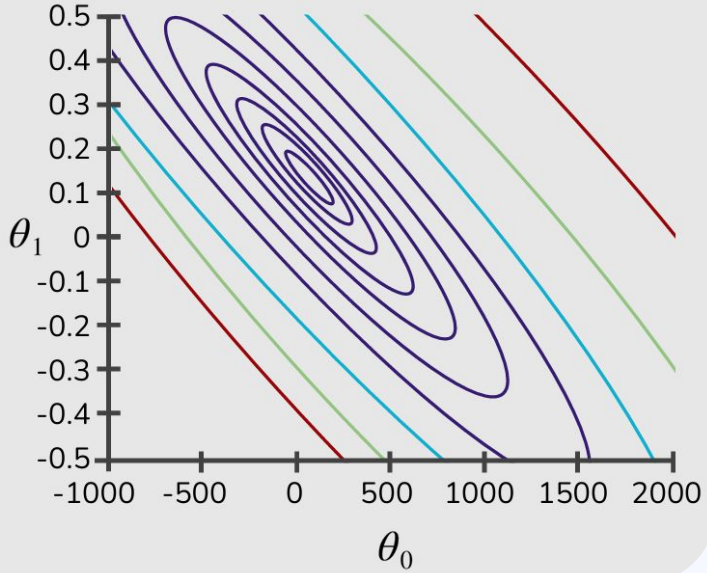
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

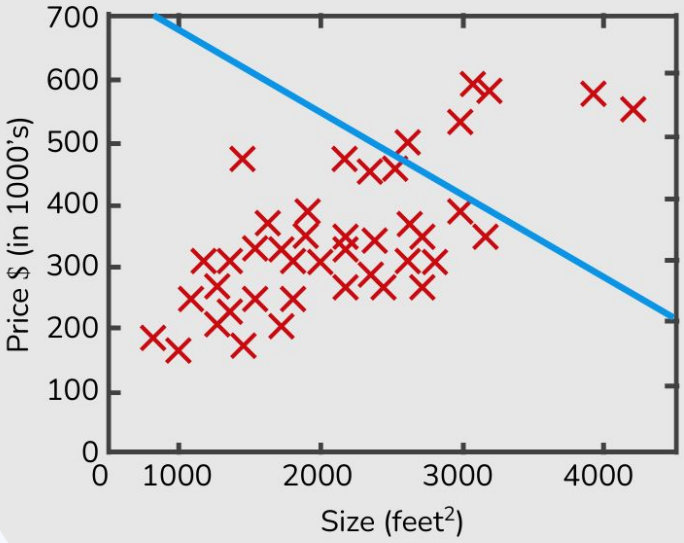
(function of the parameters θ_0, θ_1)



Função de Custo

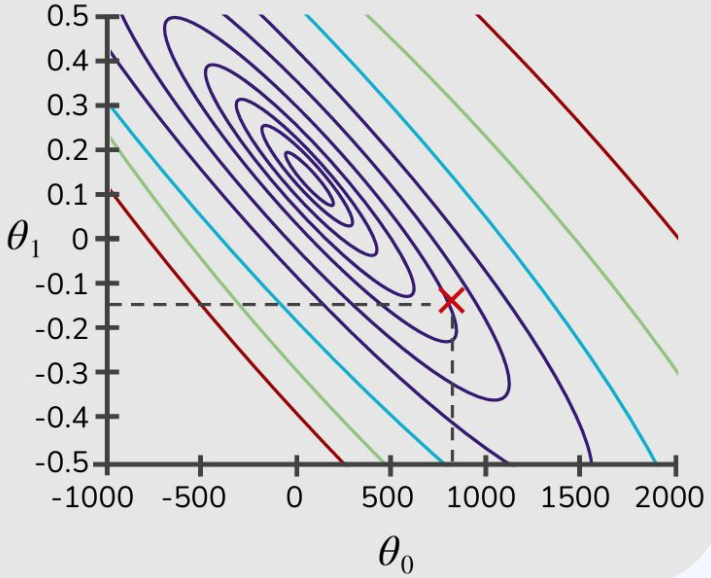
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

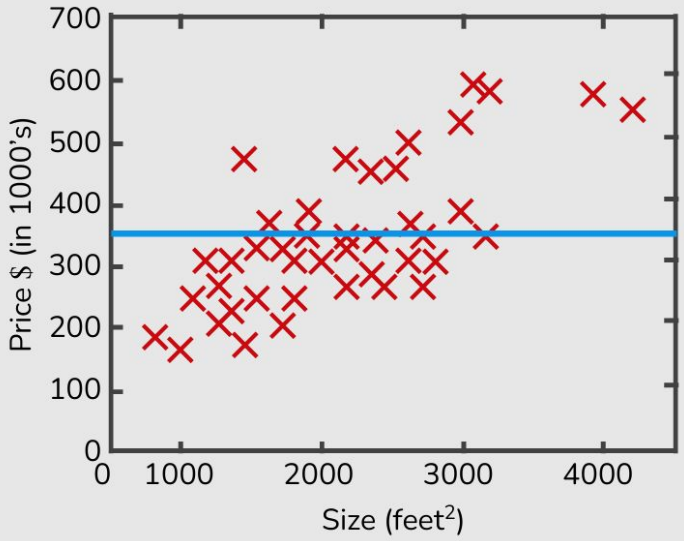
(function of the parameters θ_0, θ_1)



Função de Custo

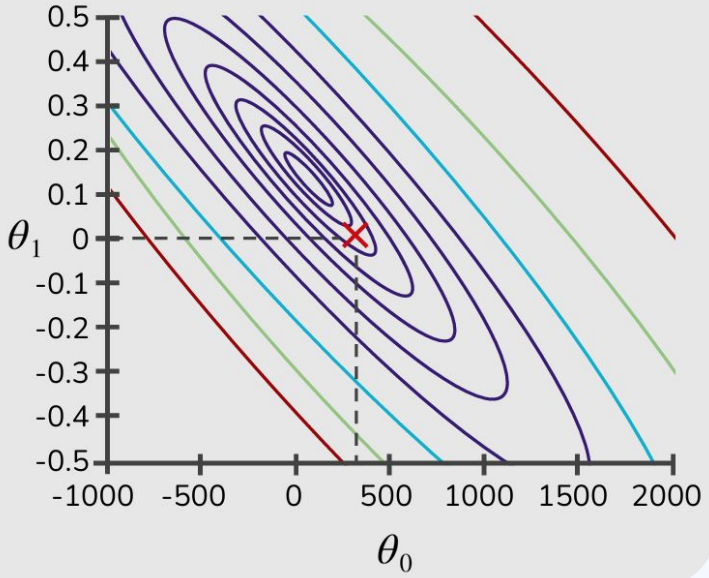
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

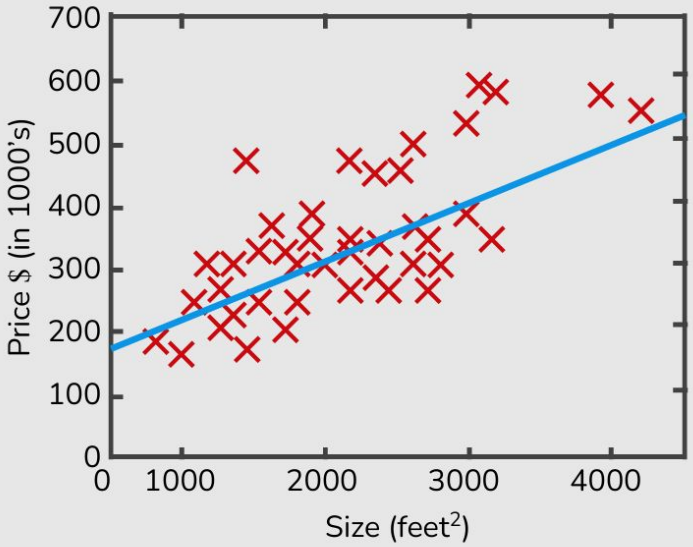
(function of the parameters θ_0, θ_1)



Função de Custo

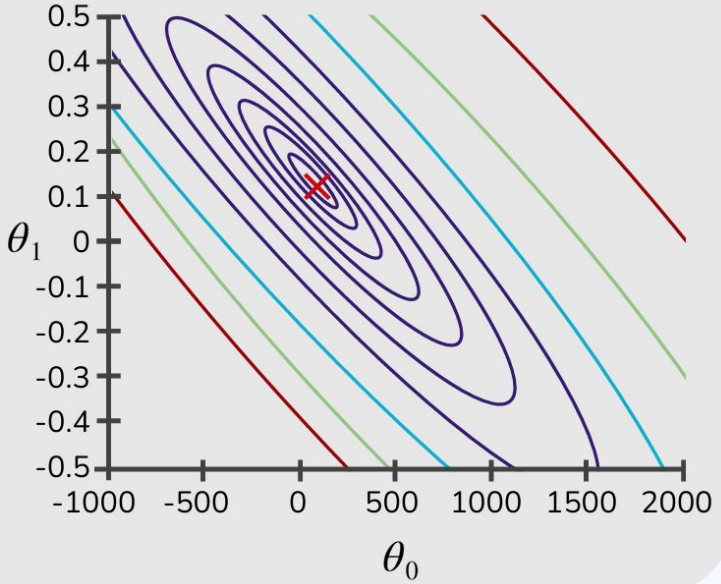
$h_\theta(x)$

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

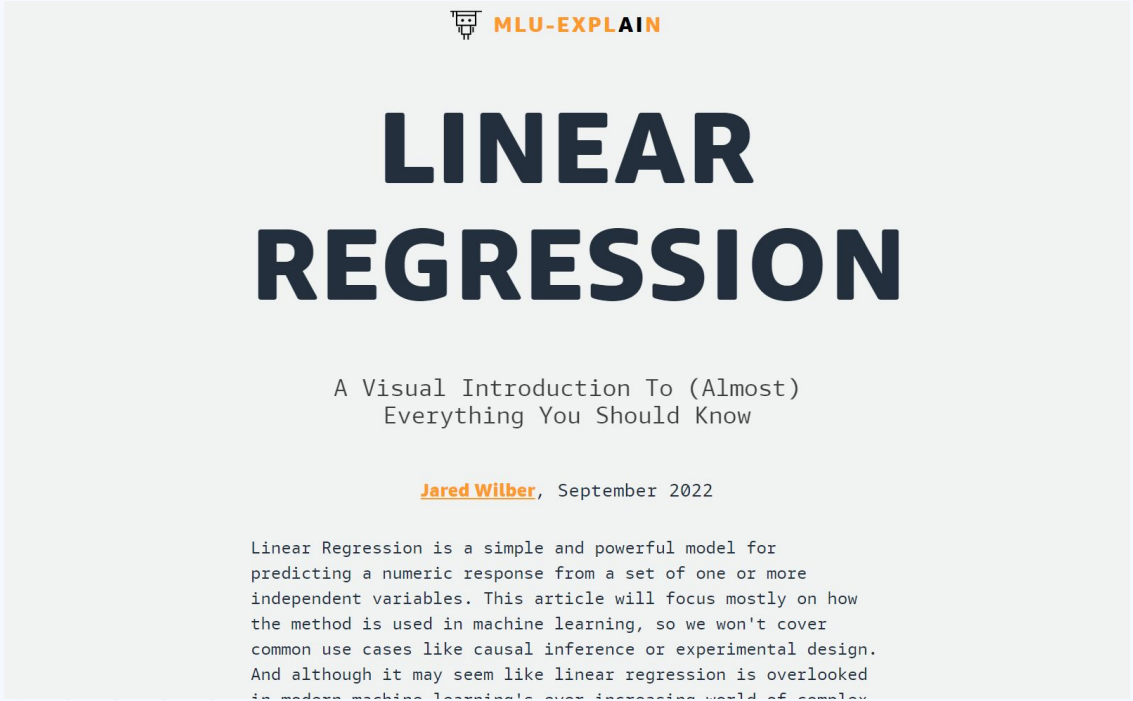


$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



A Visual Introduction To (Almost) Everything You Should Know

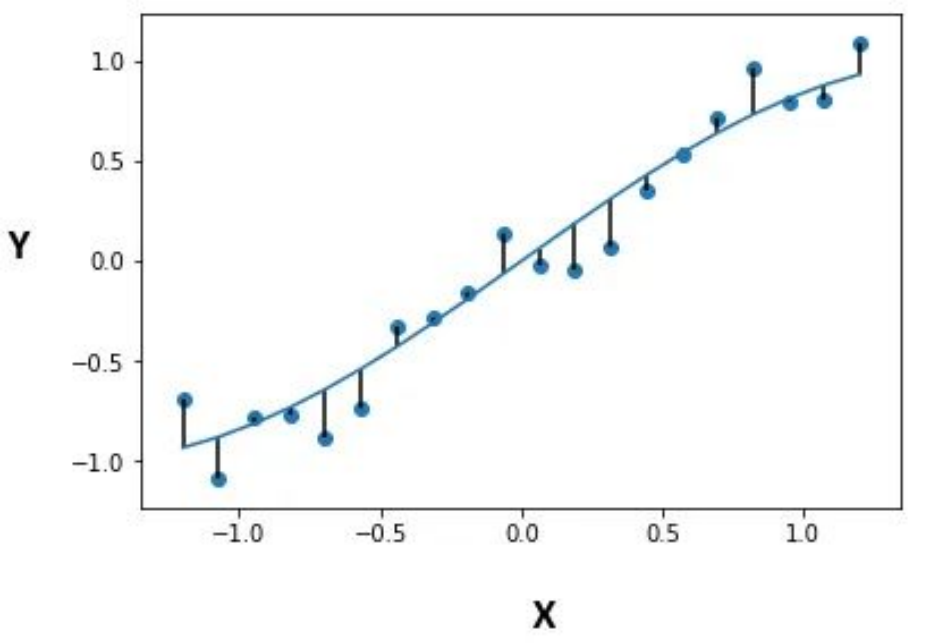


03

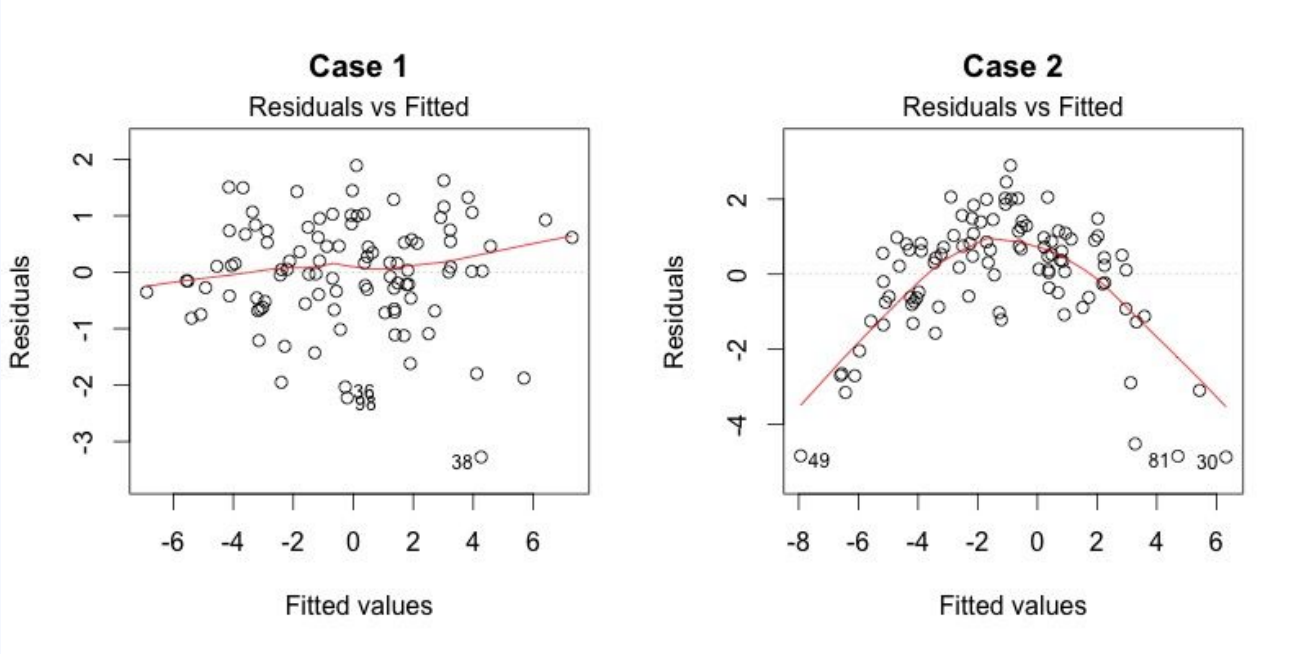
Como utilizar uma Regressão Linear?

Como saber se seus dados tem uma relação linear?

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



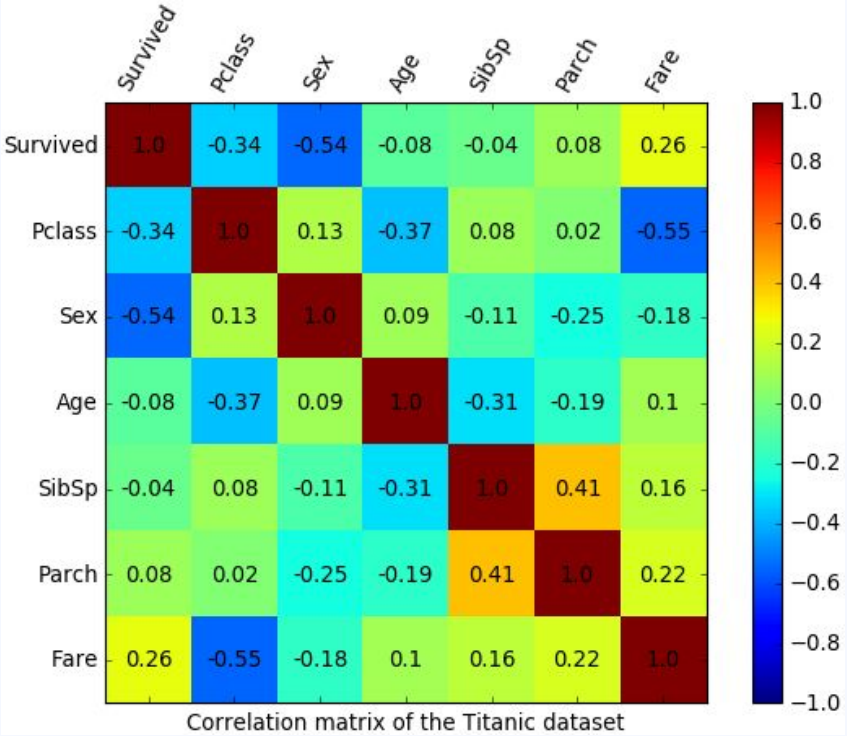
Como saber se seus dados tem uma relação linear?



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

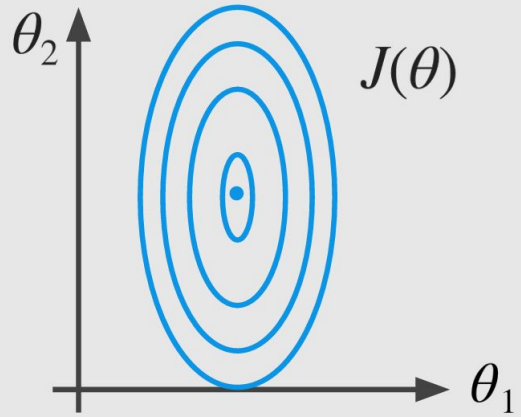
Label encoding:

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*

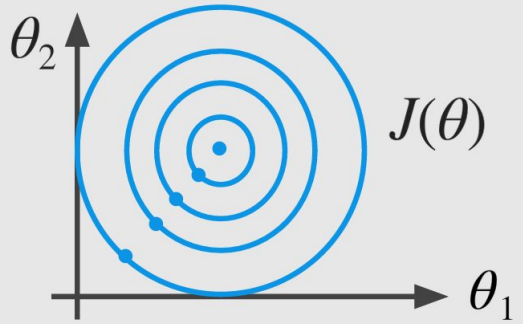
x_1 = size (0–2000 feet^2)

x_2 = number of bedrooms (1–5)



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

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

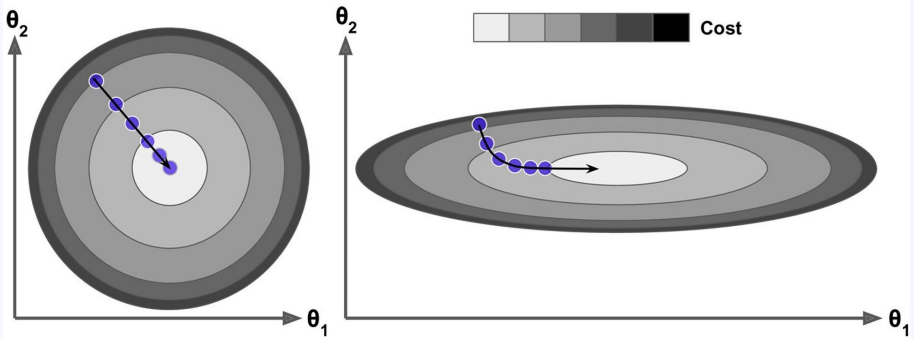


Normalizando as *features*

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

ASHISH · 5Y AGO · 174,629 VIEWS

258

Copy & Edit 3393

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 Profª. 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)

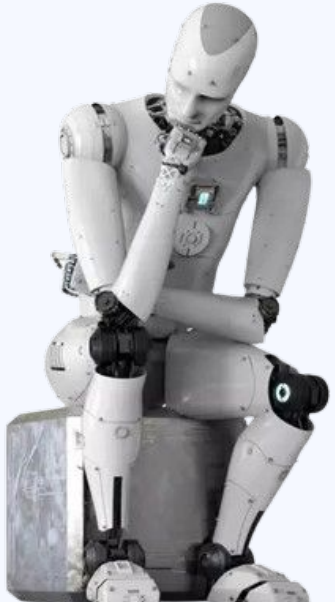
O aprendizado de máquina

Programação tradicional:

Dados + Regras → Processamento computacional → Respostas

Aprendizado de máquina:

Dados + Respostas (opcional) → Processamento computacional → Regras



Revisão e atividade prática

Ambiente de desenvolvimento: Google Colab.

Atividade de revisão:

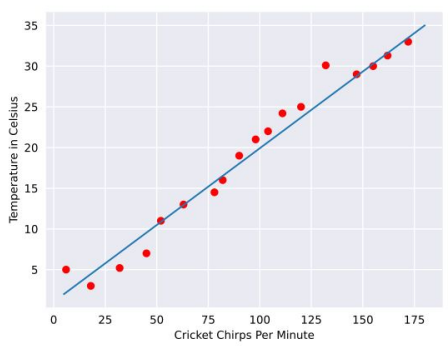
- Python;
- Manipulação de dados;
- Análise de dados;
- Operação com matrizes.

O que é Machine Learning?

"Um programa de computador **aprende a partir de uma experiência E** com respeito a uma **classe de tarefas T** e uma **métrica de performance P**, se a **sua performance em tarefas de T, quando medida por P, melhora com a experiência E**."

— **Tom Mitchell (1997)**

O que é Machine Learning?



Modelos que se ajustam aos dados:

- Regressão linear
- Regressão logística
- Algoritmos de Clusterização
- ...

Exemplos:

- Classificadores simples
- Sistemas preditivos
- Sistemas de recomendação
- ...

Retirado de: https://training.galaxyproject.org/training-material/topics/statistics/tutorials/clustering_machinelearning/tutorial.html,https://www.javatpoint.com/classification-algorithm-in-machine-learning,https://developers.google.com/machine-learning/crash-course/descending-into-ml/linear-regression

O que é Deep Learning?

"A aprendizagem profunda é uma forma de aprendizagem de máquina que permite aos computadores aprender com a experiência e **compreender o mundo em termos de uma hierarquia de conceitos**. Como o computador adquire conhecimento a partir da experiência, **não é necessário que um operador de computador humano especifique formalmente todo o conhecimento que o computador precisa**. A hierarquia de conceitos permite que o computador **aprenda conceitos complicados construindo-os a partir de conceitos mais simples**."

— Goodfellow et al. (2016)

O que é Deep Learning?

Sistemas que descobrem, aprendem e combinam padrões aprendidos:

- CNNs
- RNNs e LSTMs
- Transformers
- ...

Exemplos:

- Sistemas de visão computacional
- Chatbots modernos
- IA generativa
- ...

Retirado de:
https://exame.com/inteligencia-artificial/o-que-e-chatgpt-como-usar-a-ia-em-portugues-no-seu-dia-a-dia/
https://medium.com/analytics-vidhya/yolo-explained-5b6f4564f31,
https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/

Emergência e homogeneização

Emergência

Surgimento de novos comportamentos.
“Soma das partes não é igual ao todo”.

Homogeneização

Uso de uma mesma técnica em várias áreas.

Retirado de: https://arxiv.org/abs/2108.07258

Emergência e homogeneização

Machine Learning

Emergência	Sistemas passam a ter a capacidade de descobrirem como resolver uma tarefa a partir de dados.
Homogeneização	Diversas aplicações passaram a poder se basear em um algoritmo genérico de aprendizado com base em dados.

Deep Learning

Emergência	Aprendizagem de conceitos abstratos sem a especificação explícita. A escala enorme de modelos trouxe características como *in-context learning*.
Homogeneização	As mesmas arquiteturas de modelos passaram a ser aplicáveis para diversos tipos de aplicações.

Introduction

Mercury is the closest planet to the Sun and the smallest one in the entire Solar System. **This planet's name has nothing to do with the liquid metal,** since Mercury was named after the Roman messenger god. Mercury's surface is filled with craters

Mercury takes a little more than 58 days to complete its rotation, so try to imagine how long days must be there! **Since the temperatures are so extreme, albeit not as extreme** as on Venus, Mercury has been deemed to be non-habitable for humans

“This is a quote, words full of wisdom that someone important said and can make the reader get inspired”

—Someone Famous

Concepts

Mercury

Mercury is the closest planet to the Sun and **the smallest one** in the Solar System—it's only a bit larger than the Moon

Venus

Venus has a beautiful name and is the **second planet from the Sun.** It's hot and has a poisonous atmosphere

What is this topic about?

Mercury

It’s the closest planet to the Sun and the **smallest** in the Solar System

Venus

Venus has a beautiful name and is the second planet from the Sun

Mars

Despite being red, Mars is actually a **cold place.** It’s full of iron oxide dust

Features of the topic

Mars

Despite **being red**, Mars is very cold

Neptune

It’s the farthest planet from the Sun

Jupiter

Jupiter is the biggest planet of them all

Saturn

Saturn is a **gas giant** and has several rings

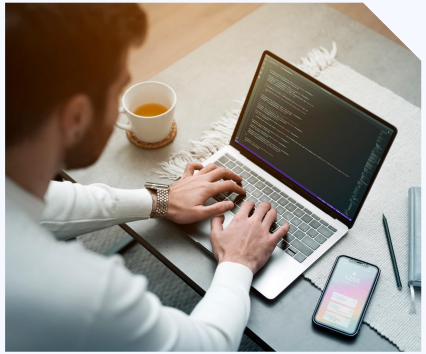
Examples

Mercury

It’s the closest planet to the Sun and the **smallest** in the Solar System

Venus

Venus has a beautiful name and is the second planet from the Sun



Mars

Despite being red, Mars is actually a **cold place.** It’s full of iron oxide dust

Recommendations

Mars

Despite being red, Mars is very cold

Venus

Venus is the second planet from the Sun

Neptune

Neptune is the farthest planet from the Sun

Mercury

Mercury is the closest planet to the Sun

Saturn

Saturn is a gas giant and has several rings

Jupiter

Jupiter is the biggest planet of them all

Image always reinforce the concept

You can give a brief description of the topic you want to talk about here. For example, if you want to talk about Mercury, you can say that it’s the smallest planet in the entire Solar System

4,498,300,000

Big numbers catch your audience’s attention

9h 55m 23s

Jupiter’s rotation period

333,000

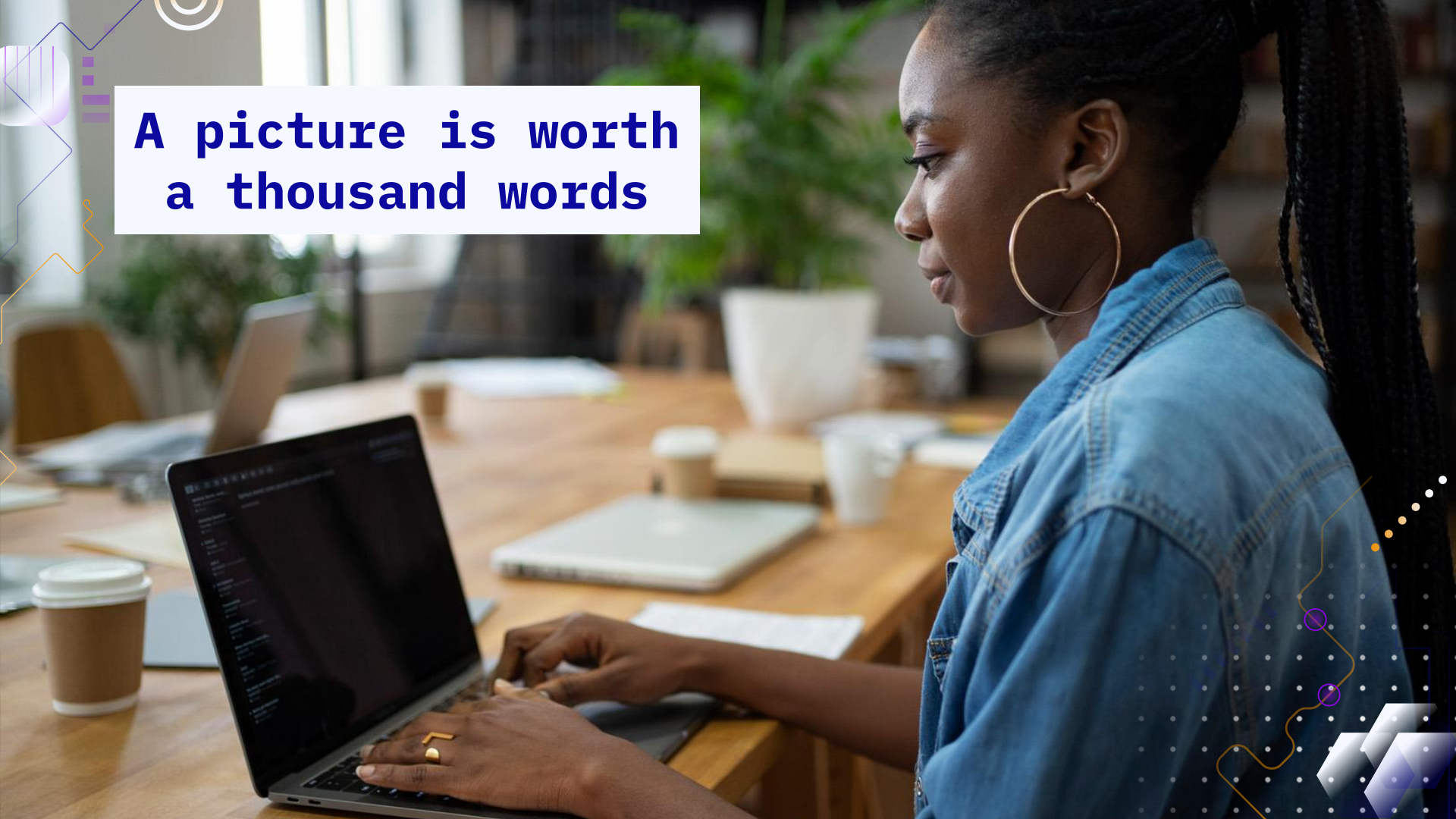
The Sun’s mass compared to Earth’s

386,000 km

Distance between the Earth and the Moon

Awesome words

A picture is worth a thousand words



Practical exercise - calculator

Objective:

Introduce participants to basic coding concepts by building a **simple calculator**

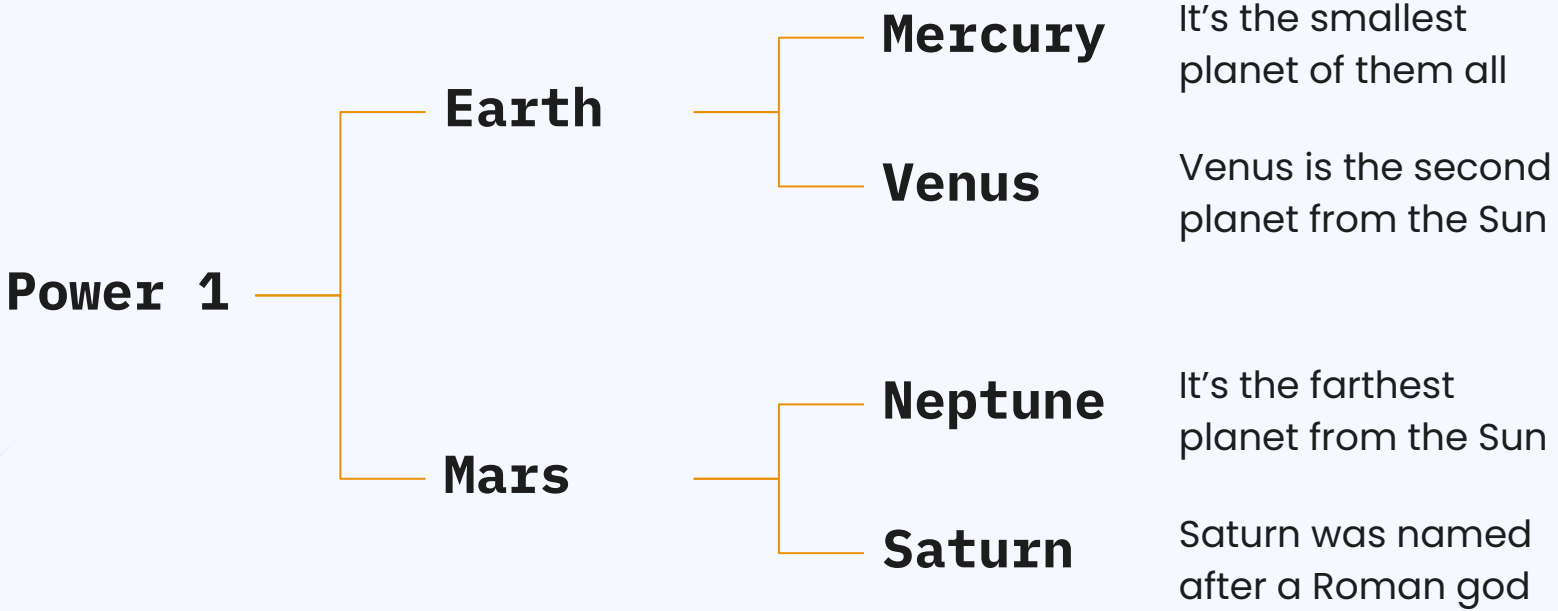
Instructions:

1. Open a Python development environment and write the following code:

```
# Simple Calculator
num1 = int(input("Enter the first number: "))
num2 = int(input("Enter the second number: "))
print("Sum:", num1 + num2)
print("Difference:", num1 - num2)
print("Product:", num1 * num2)
print("Quotient:", num1 / num2)
```

2. Run the program and experiment with different numbers
3. Observe the output

Brainstorm and idea generation



Main topic and details

Mars

Despite being red, Mars is **very cold**

Neptune

It’s the farthest planet from the Sun

Jupiter

Jupiter is the biggest planet of them all

Saturn

It’s a gas giant and has **several rings**

Popular programming languages

01 — **Neptune** — Mercury is the closest planet to the Sun and the **smallest** of them all

02 — **Venus** — Venus has a beautiful name and is the **second planet from the Sun**

03 — **Earth** — Earth is the third planet from the Sun and the only one that harbors life in the Solar System

04 — **Saturn** — Saturn is a gas giant and has several rings. It's composed mostly of hydrogen and helium

Sequences

Saturn is composed of **hydrogen and helium**

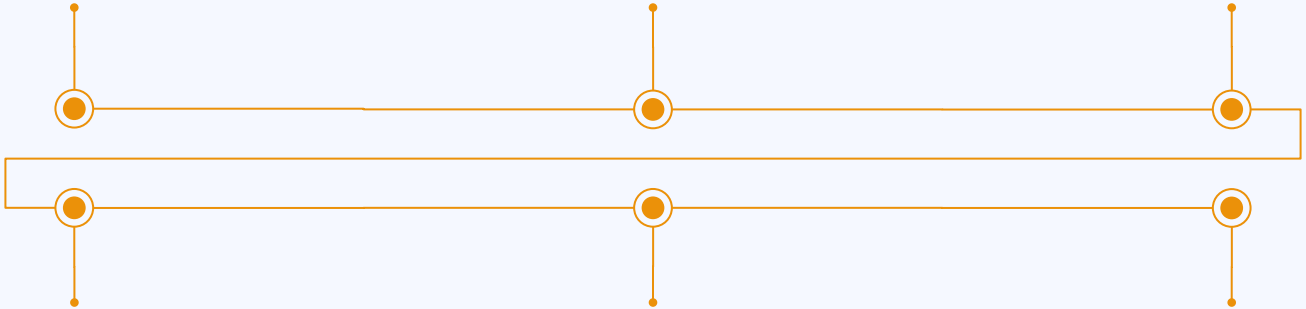
First

Mercury is the **closest** planet to the Sun

Next

Jupiter was named after a Roman god

Next



Despite being red, Mars is **very cold**

Next

Earth is the third planet from the Sun

Next

Venus has extremely high temperatures

Last

Classification

Mars	Venus	Mercury	Jupiter
• Small • Red • Cold • Rocky	• Small • Hot • Dry • Volcanic	• Small • Hot • Rocky • Cratered	• Large • Cold • Gassy • Striped
Mars is full of iron oxide dust	Venus has high temperatures	Mercury is quite a small planet	Jupiter is a huge gas giant

Cause and effect

Problem

Mars

Despite being red, Mars is very cold

Venus

Venus is the second planet from the Sun

Solution

Mercury

Mercury is the closest planet to the Sun

Saturn

Saturn is a gas giant and has several rings

Question and answer

Question

Is Mercury the closest planet to the Sun and the smallest one in the Solar System? **Note that it’s a bit larger than the Moon**

Answer

Venus has a beautiful name and is **the second planet from the Sun.** It’s hot and has a poisonous atmosphere

Step-by-step coding

01

Earth

It’s the only planet known to **harbor life**

02

Mercury

Mercury is the closest planet to the Sun

03

Jupiter

Jupiter is the **biggest** planet of them all

04

Saturn

Saturn was named after a Roman god

Parts and whole

The whole objective

Mercury is the closest planet to the Sun and the smallest one in the entire Solar System

Parts of the object

- Mercury
- Jupiter
- Venus
- Mars
- Earth
- Saturn
- Mercury

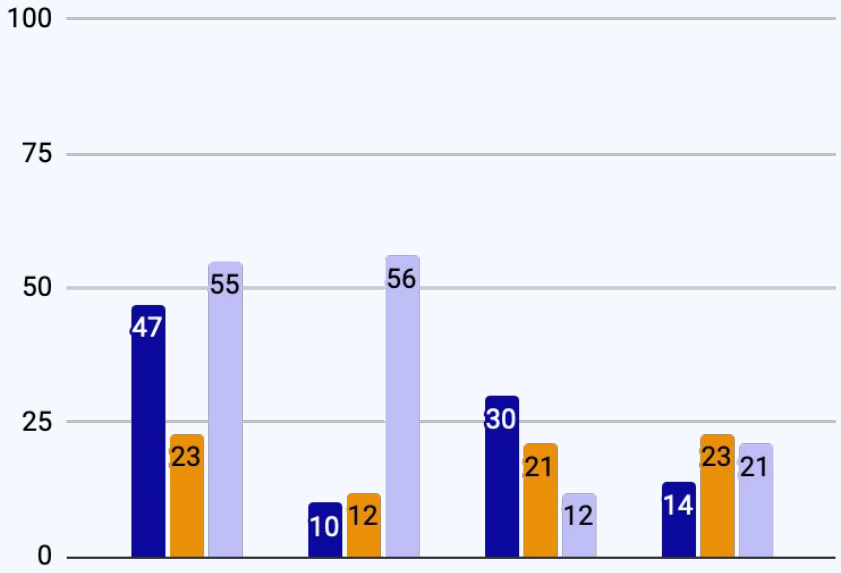
What happens if the parts are missing?

Earth is the third planet from the Sun and the **only one that harbors life in the Solar System**

What's the function of the parts?

Jupiter is a gas giant and the biggest planet in the Solar System

You can use this graph



Follow the link in the graph to modify its data and then paste the new one here. **For more info, click here**

Mercury

Mercury is the closest planet to the Sun

Jupiter

Jupiter is the biggest planet of them all

Saturn

Saturn was named after a Roman god

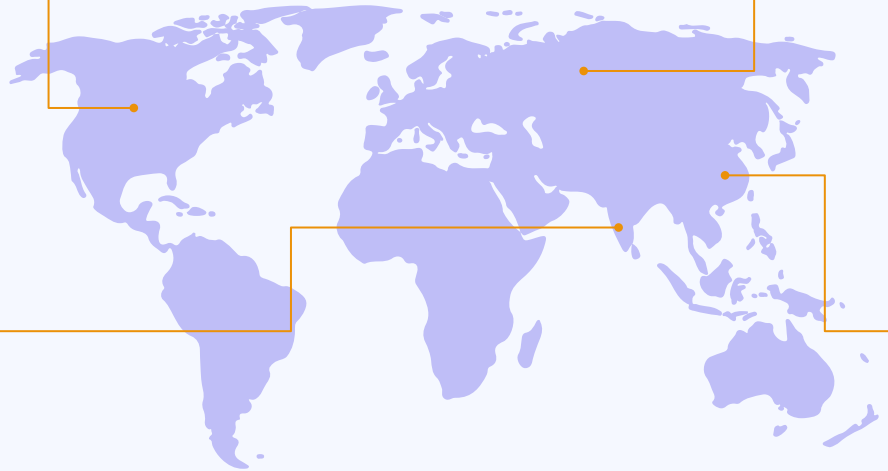
This is a map

USA

Despite being red, Mars is **very cold**

India

Jupiter is the biggest planet of them all



Russia

Neptune is the farthest planet from the Sun

China

Saturn is **a gas giant** and has several rings

Mockups

You can replace the images on the screen with your own work. Just right-click on them and select “Replace image”

Obrigado!

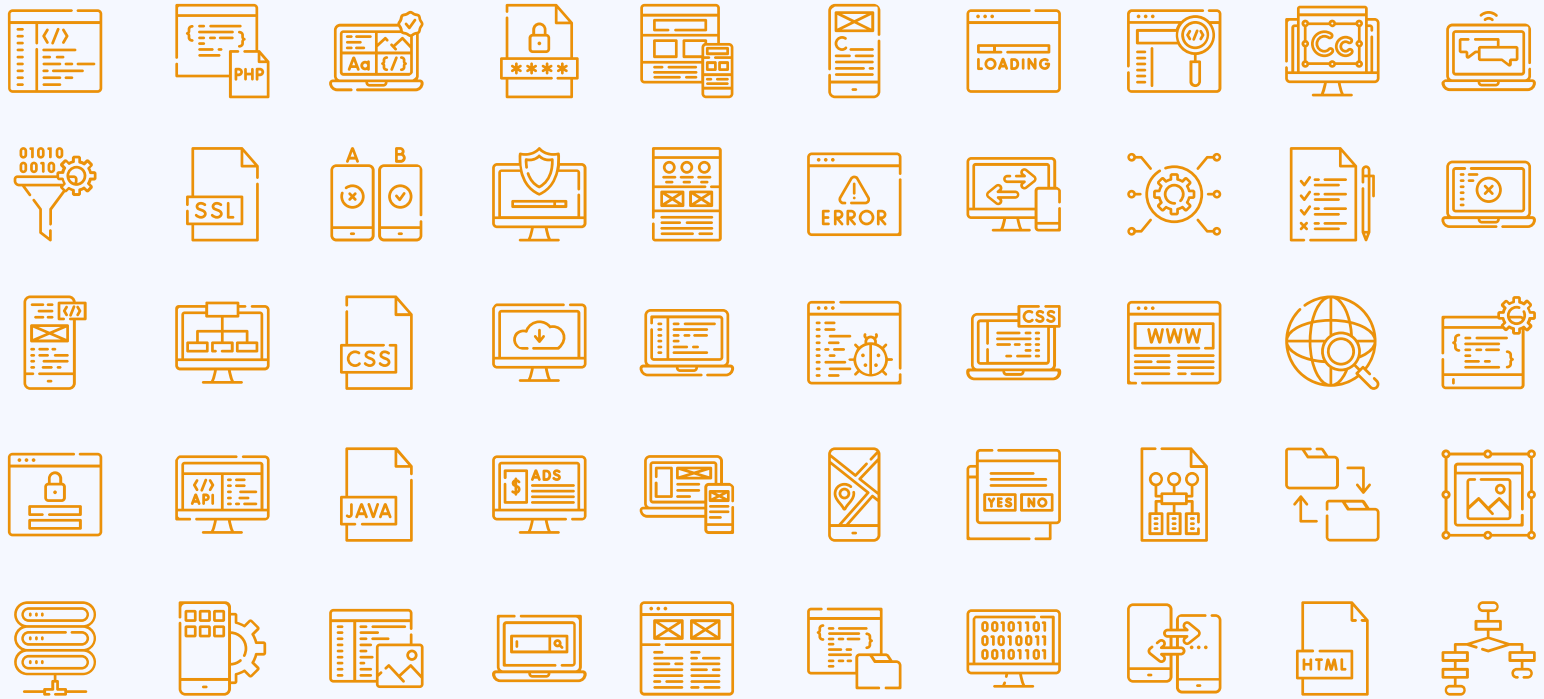
Do you have any questions?

@irisdatascienceunicamp



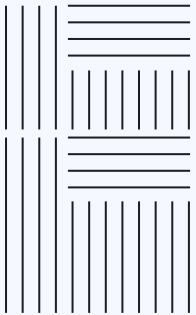
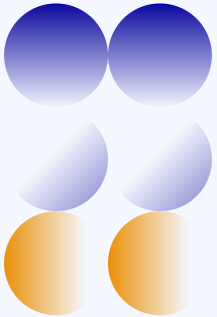
CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Icon pack


PHP
****
LOADING
01010
0010
SSL
A
B
ERROR
CSS
WWW
ADS
API
JAVA
YES
NO
HTML

Alternative resources

Here’s an assortment of alternative resources whose style fits that of this template:

Resources

Did you like the resources in this template? Get them for free at our other websites:

Photos:

- Medium shot man working on laptop
- Group of friends planning a trip in a cafe
- Medium shot man working on laptop
- Side view of men working on laptops at the office
- Lifestyle of woman in the office
- Secretary working on laptop

Vectors:

- Abstract gradient circuit board background

Icon pack:

- Icon Pack: Coding | Lineal

Agenda

01

Apresentação do curso

You can describe the topic of the section here

02

O que é IA, ML e DL?

You can describe the topic of the section here

03

Tips

You can describe the topic of the section here

04

Problemáticas atuais

You can describe the topic of the section here

Contents of this template

You can delete this slide when you're done editing the presentation

Fonts	To view this template correctly in PowerPoint, download and install the fonts we used
Used and alternative resources	An assortment of graphic resources that are suitable for use in this presentation
Thanks slide	You must keep it so that proper credits for our design are given
Colors	All the colors used in this presentation
Icons and infographic resources	These can be used in the template, and their size and color can be edited
Editable presentation theme	You can edit the master slides easily. For more info, click **here**

For more info:
Slidesgo | **Blog** | **FAQs**

You can visit our sister projects:
Freepik | **Flaticon** | **Storyset** | **Wepik** | **Videvo**

Instructions for use

If you have a free account, in order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide. Please refer to the next slide to read the instructions for premium users.

As a Free user, you are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit our blog:
https://slidesgo.com/faqs and **https://slidesgo.com/slidesgo-school**

Instructions for use (premium users)

As a Premium user, you can use this template without attributing **Slidesgo** or keeping the **Thanks** slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the “Thanks” slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

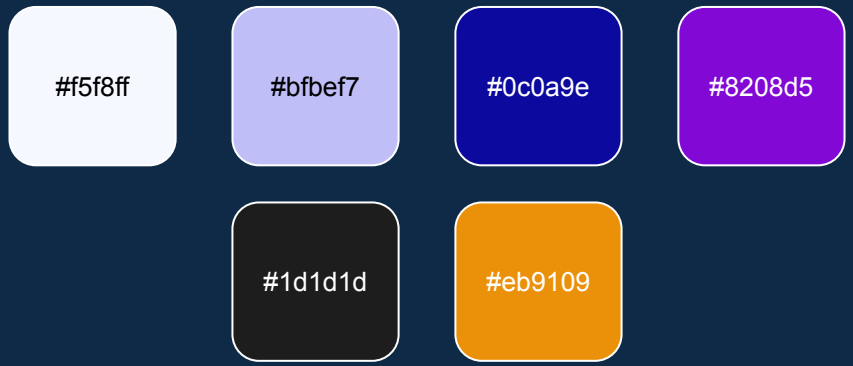
For more information about editing slides, please read our FAQs or visit our blog:
https://slidesgo.com/faqs and **https://slidesgo.com/slidesgo-school**

Fonts & colors used

This presentation has been made using the following fonts:

IBM Plex Mono Bold
(https://fonts.google.com/specimen/IBM+Plex+Mono)

Poppins Normal & Bold
(https://fonts.google.com/specimen/Poppins)



Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out **how it works**.

Pana

Amico

Bro

Rafiki

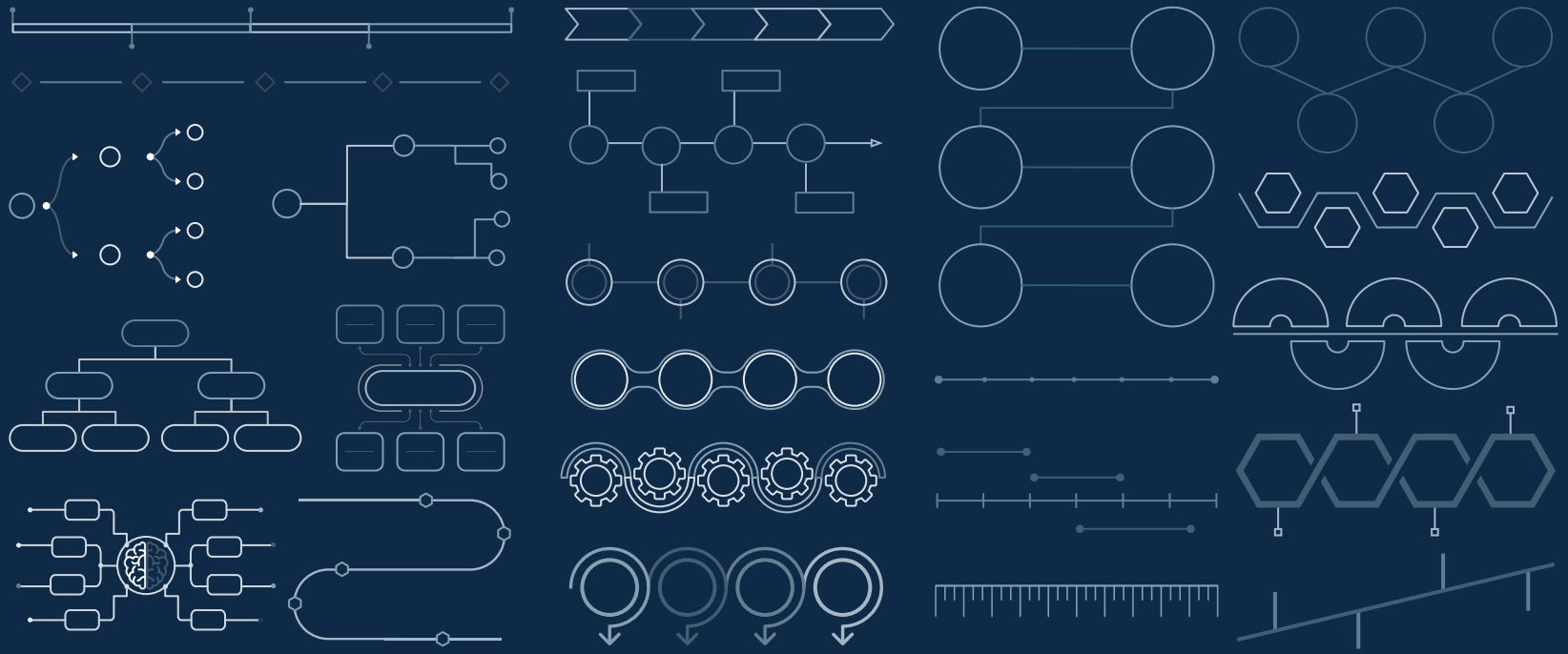
Cuate

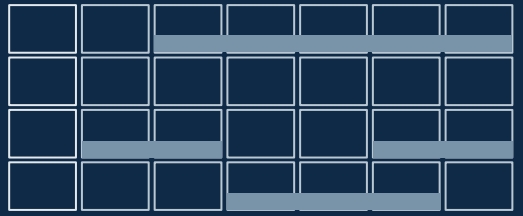
Use our editable graphic resources...

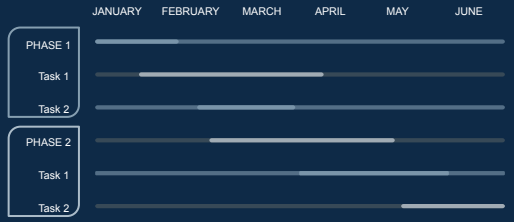
You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.

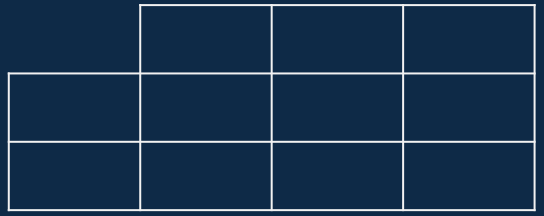


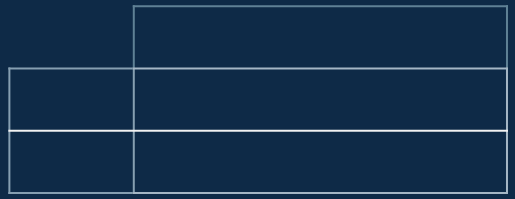


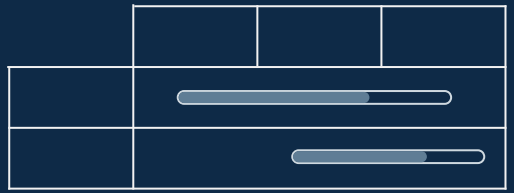



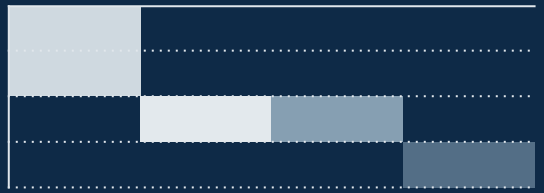


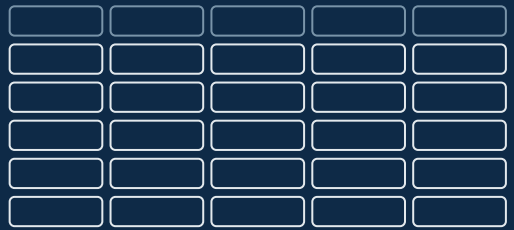

JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2

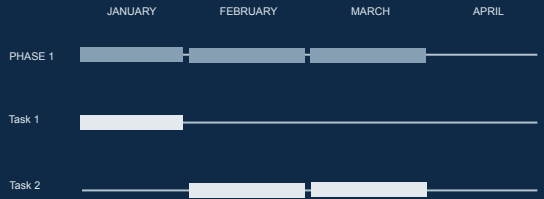


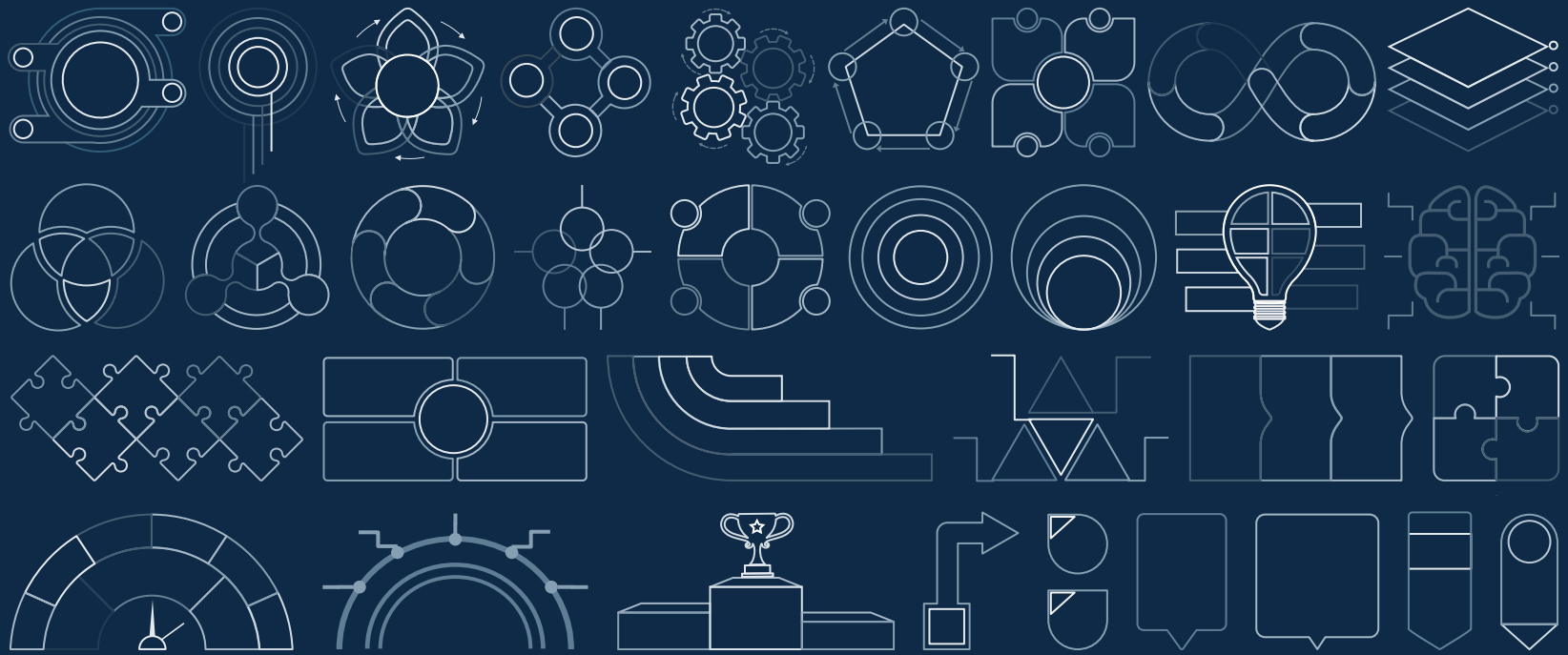


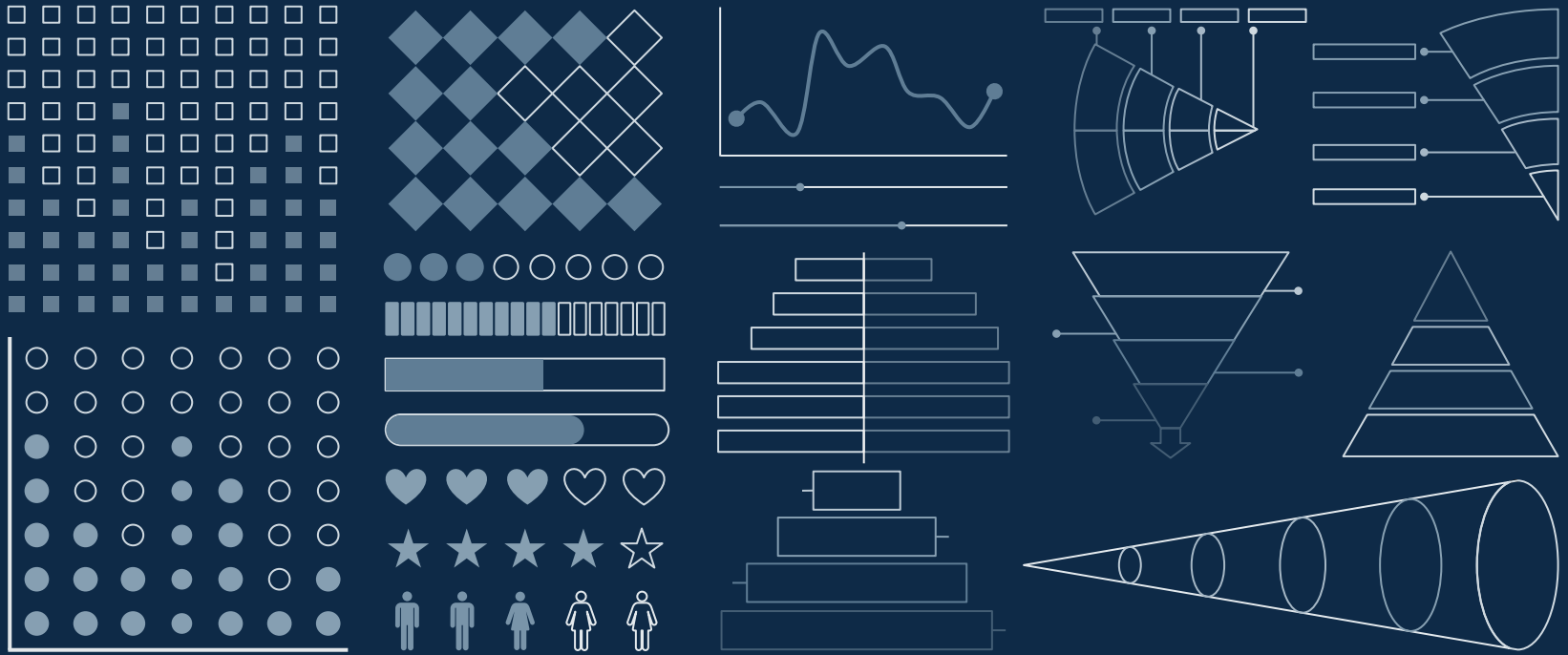







JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2





...and our sets of editable icons

You can **resize** these icons without losing quality.
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.
In Google Slides, you can also use **Flaticon's extension**, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons


slidesgo