

INTELIGÊNCIA ARTIFICIAL & BIG DATA

Profª . Miguel Bozer da Silva

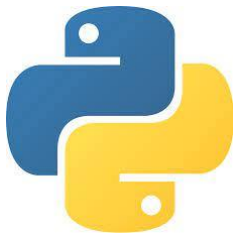
Tópicos da disciplina



- Revisão de ciência de dados;
- Aprendizado Supervisionado e não Supervisionado;
- Redução de Dimensionalidade;
- Métricas de Desempenho;
- Redes Neurais Artificiais;

Ferramentas para a disciplina

- Nossas aulas serão muito HANDS ON
- Iremos utilizar a linguagem em python para utilizarmos bibliotecas de IA.

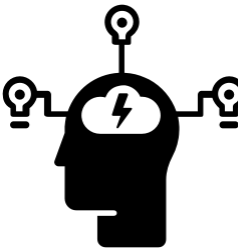


Afinal, o que é Inteligência Artificial



Definição de Inteligência Artificial

- Ainda não existe uma boa definição pois ainda não sabemos o que realmente é inteligência!
“Inteligência é a capacidade de **analisar** uma determinada situação, **tomar** uma **decisão** e **aprender** através da **compreensão** do resultado”.
- Entretanto, podemos esboçar algumas ideias:
“IA são sistemas artificiais que aparentam possuir inteligência”.
“IA são sistemas artificiais que realizam atividades semelhantes as atividades cognitivas humanas”.
- **Importante:** um sistema apresentar Inteligência Artificial não significa que o sistema possui consciência! **IA ≠ consciência**
- E também ainda não sabemos o que é consciência!



Definição de Inteligência Artificial

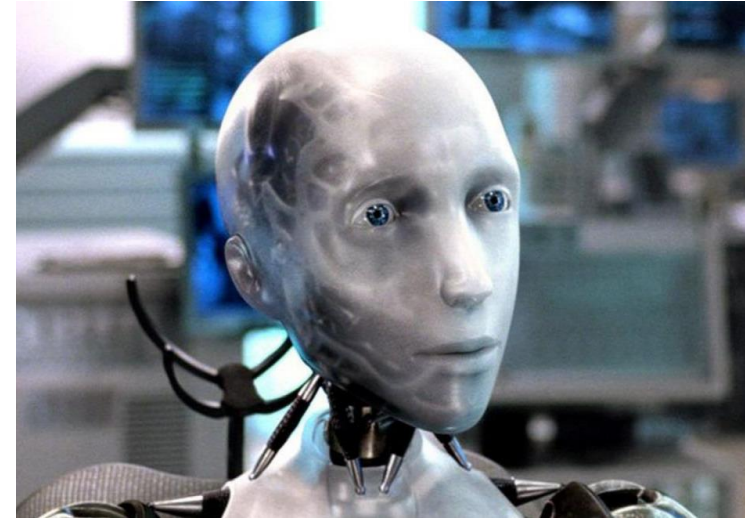
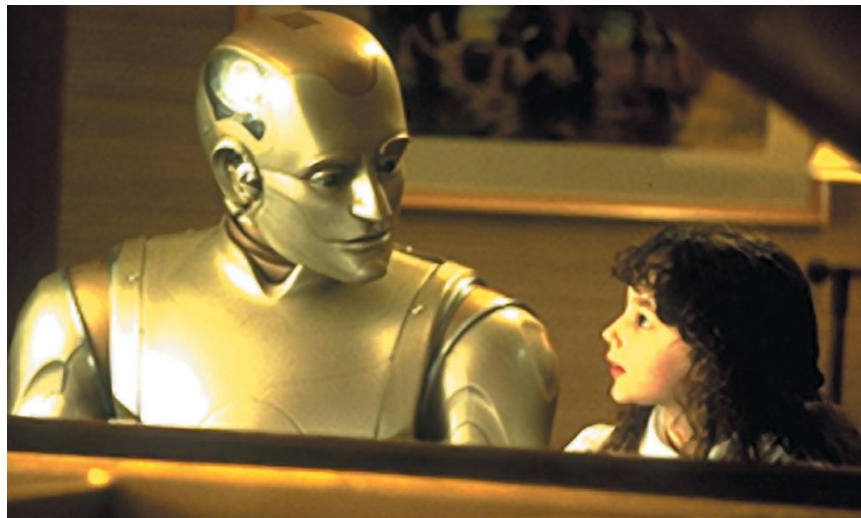
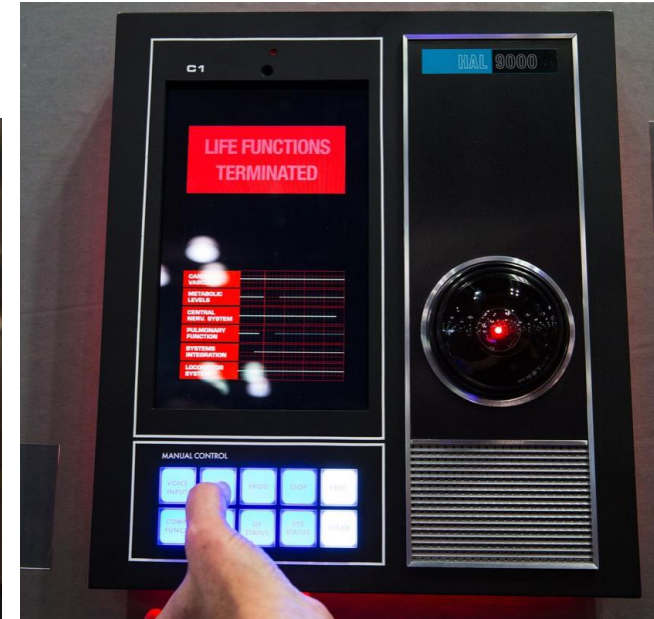


Existem dois tipos de IA:

- 1) Inteligência artificial de **propósito geral**: um sistema que consegue modificar sua função e propósito, atuando em ambientes complexos com muitas variáveis e informação incompleta -> NUNCA FOI FEITO; atualmente existem poucos esforços de pesquisa nesse sentido.
- 2) Inteligência artificial de **propósito específico**: um sistema feito para atuar sobre um problema específico, conseguindo lidar com variáveis conhecidas ou que podem ser obtidas ou inferidas. Atualmente existem muitas pesquisas e muitas aplicações de mercado desses sistemas!

Nesse curso vamos estudar apenas as de **propósito específico**.

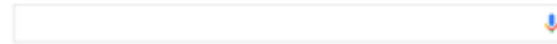
IA de propósito geral



IA de propósito específico



Google



Buscar con Google

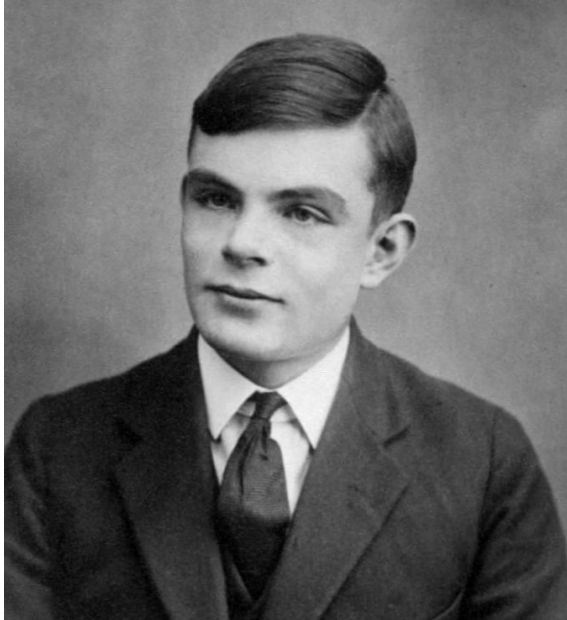
Me siento con suerte



facebook



IA de propósito específico: interação humano-computador (primórdios)



Alan Turing (1912-1954), um dos pais da computação

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.

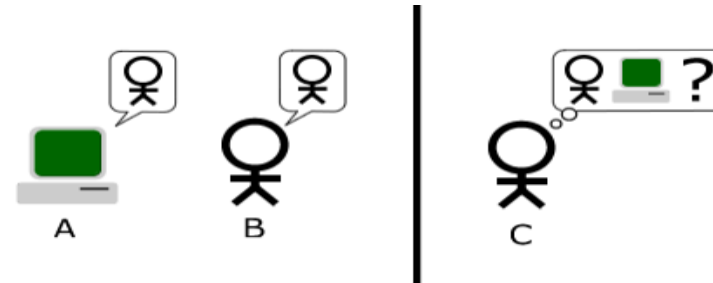
COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be

Teste de Turing



IA de propósito específico: interação humano-computador (primórdios)



```
Welcome to

EEEEEE LL      IIII ZZZZZZ AAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LL      II      ZZ  AAAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LLLLLL IIII ZZZZZZ AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

```
Did you come to me because you are not ?

>i dont know
What does that suggest to you?

>that i am unintelligent
How long have you been unintelligent?

>since i was born
I see.

>arent you supposed to cheer me up
You're not really talking about me, are you?

>no im talking about me
Why not?

>mrh
I'm not sure I understand you fully.

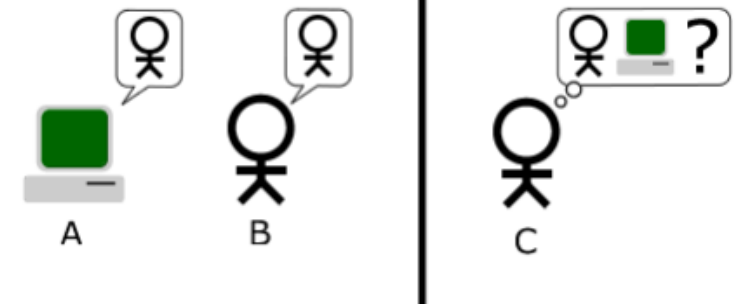
>rg
Come come elucidate your thoughts.

>what is elucidate
What do you think?

>to see
Can you elaborate on that?

>
```

Teste de Turing



Primeiro chatbot feito entre 1964 e 1966 pelo cientista da computação Joseph Weizenbaum

Um dos primeiros processadores de linguagem natural que rodava um script de interação chamado DOCTOR

IA de propósito específico: interação humano-computador (hoje no mercado)



Robôs de Companhia



Sony AIBO,
1999

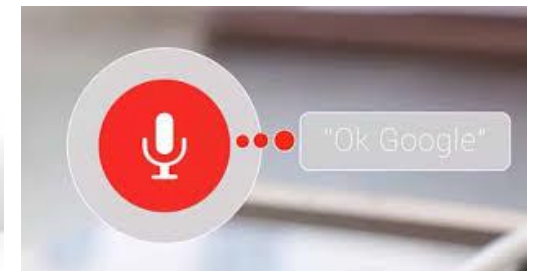


NÃO,
2006



Pepper, 2014

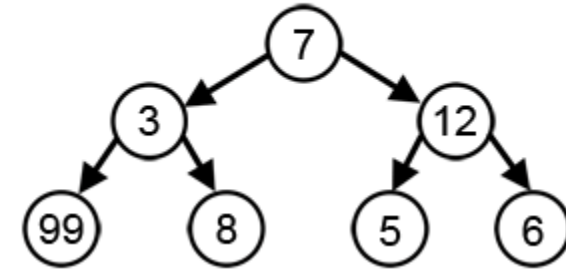
Assistentes pessoais



IA de propósito específico: tipos

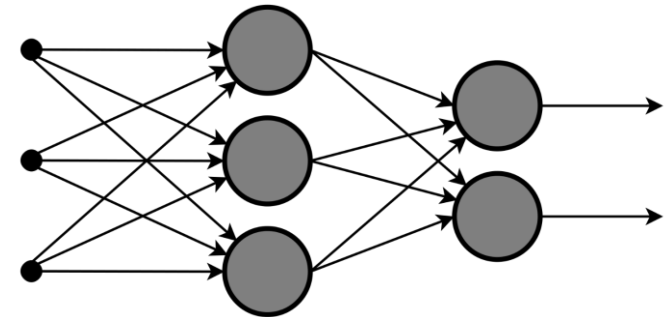
❑ **Agentes Racionais**: procedimentos puramente matemáticos que tentam solucionar o problema. Exemplos:

- Busca Gulosa e Busca A*
- Algoritmo MinMax



❑ **Sistemas bioinspirados**: procedimentos matemáticos que imitam sistemas biológicos como redes neurais e evolução. Exemplos:

- Redes Neurais Artificiais e Deep Learning
- Algoritmos Genéticos

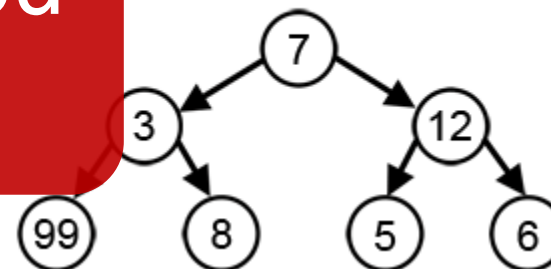


IA de propósito específico: tipos

- ❑ **Agentes Racionais:** procedimentos puramente matemáticos que tentam solucionar o problema. Exemplos:

- Busca Gulosa e Busca A*
- Algoritmo MinMax

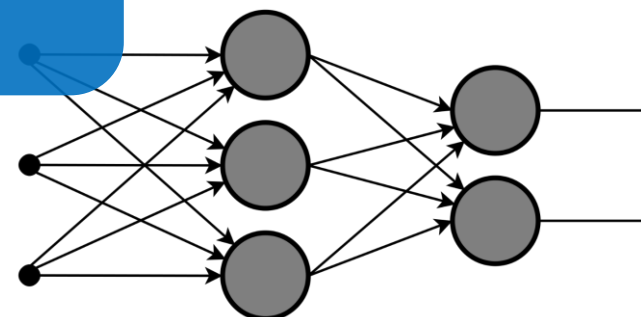
IA convencional ou
IA clássica



- ❑ **Sistemas bioinspirados:** procedimentos matemáticos que imitam sistemas biológicos como redes neurais e evolução. Exemplos:

- Redes Neurais Artificiais e Deep Learning
- Algoritmos Genéticos

Aprendizado de
Máquina

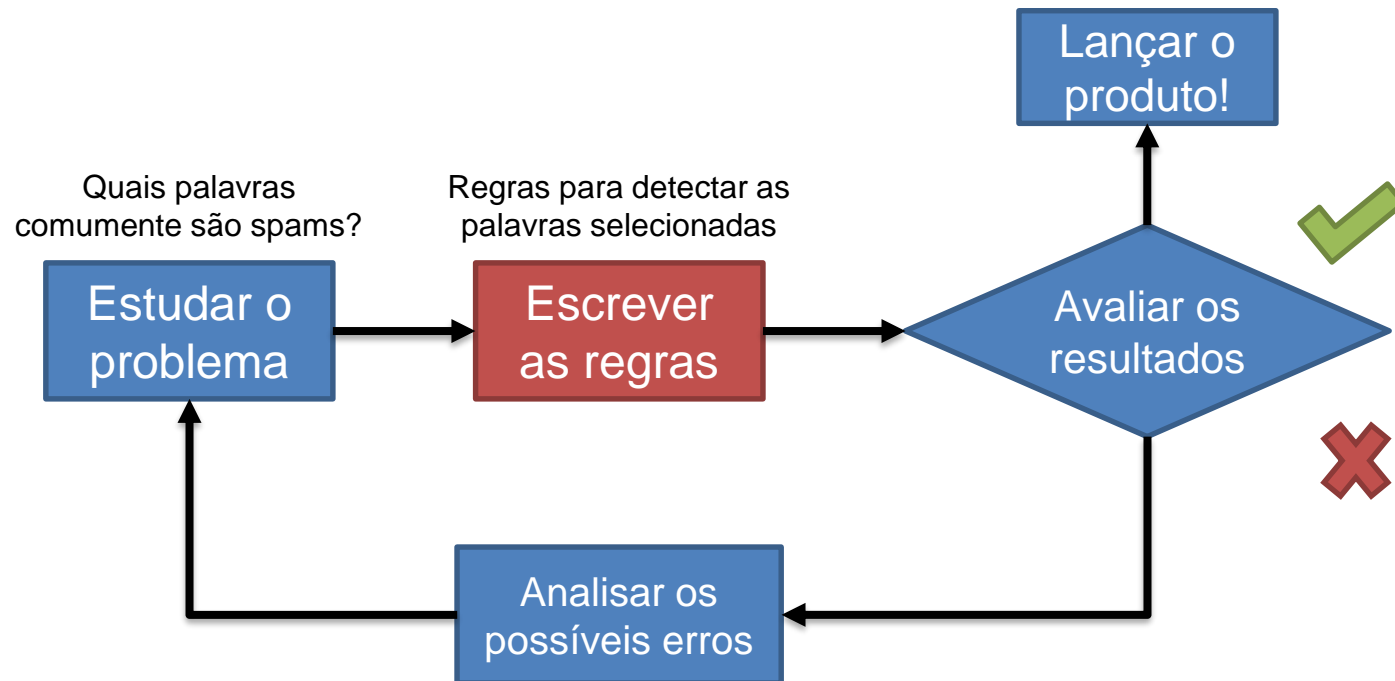


O que é o Aprendizado de Máquina?

- O que é o Aprendizado de Máquina?
 - Aprendizado de máquina ou *Machine Learning* é a ciência ou a arte de programar computadores para que eles consigam aprender dado um **conjunto de dados** (GÉRON, 2017)
 - Para ilustrar essa ideia, vamos imaginar que desejamos **criar um programa que detecte spams** na nossa caixa de e-mail

Introdução à IA

- O que é o Aprendizado de Máquina?
- Numa abordagem tradicional a forma de resolver esse problema seria a seguinte:

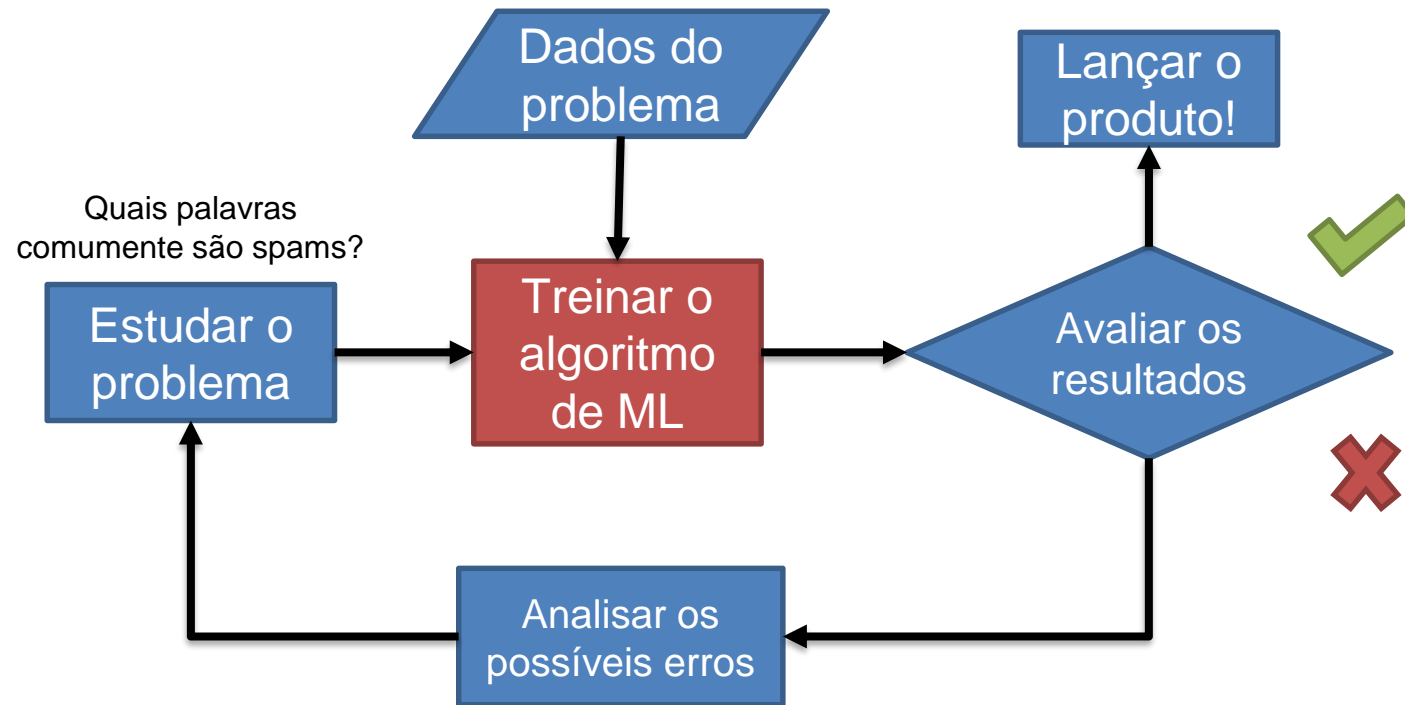


Adaptado de (GÉRON, 2017)

- O que é o Aprendizado de Máquina?
 - Como o problema **não é trivial**, seu programa poderia ter um conjunto complexo de regras;
 - Muita energia do programador para criar a solução e provavelmente um programa complexo para ser utilizado.

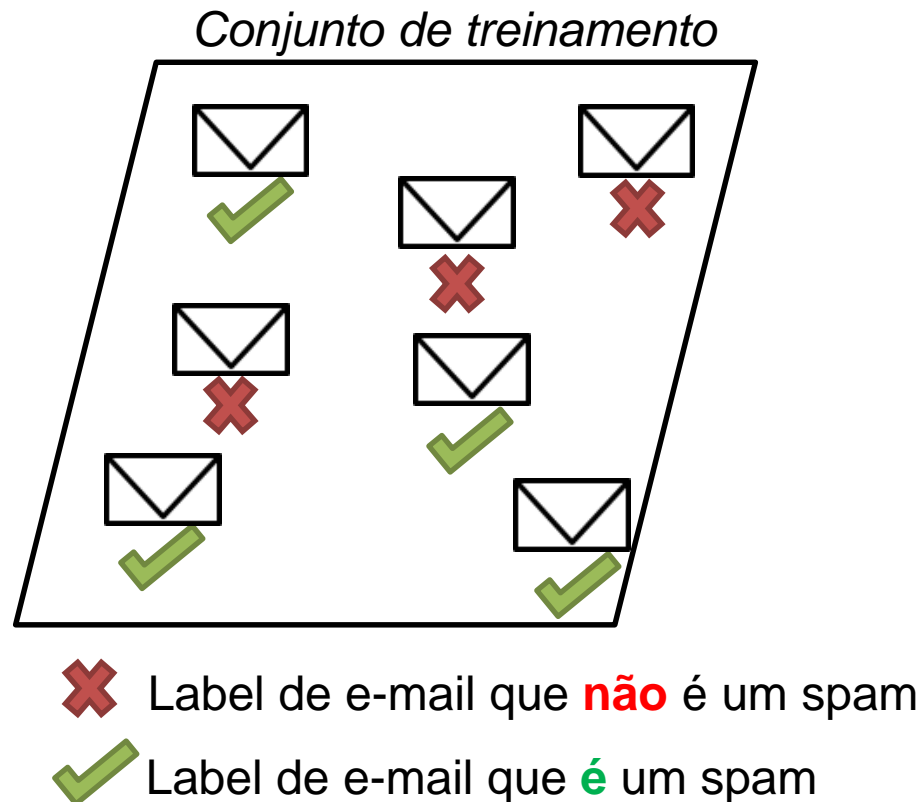
Introdução à IA

- O que é o Aprendizado de Máquina?
- Numa abordagem com ML o algoritmos irá aprender sobre o problema:



Introdução à IA

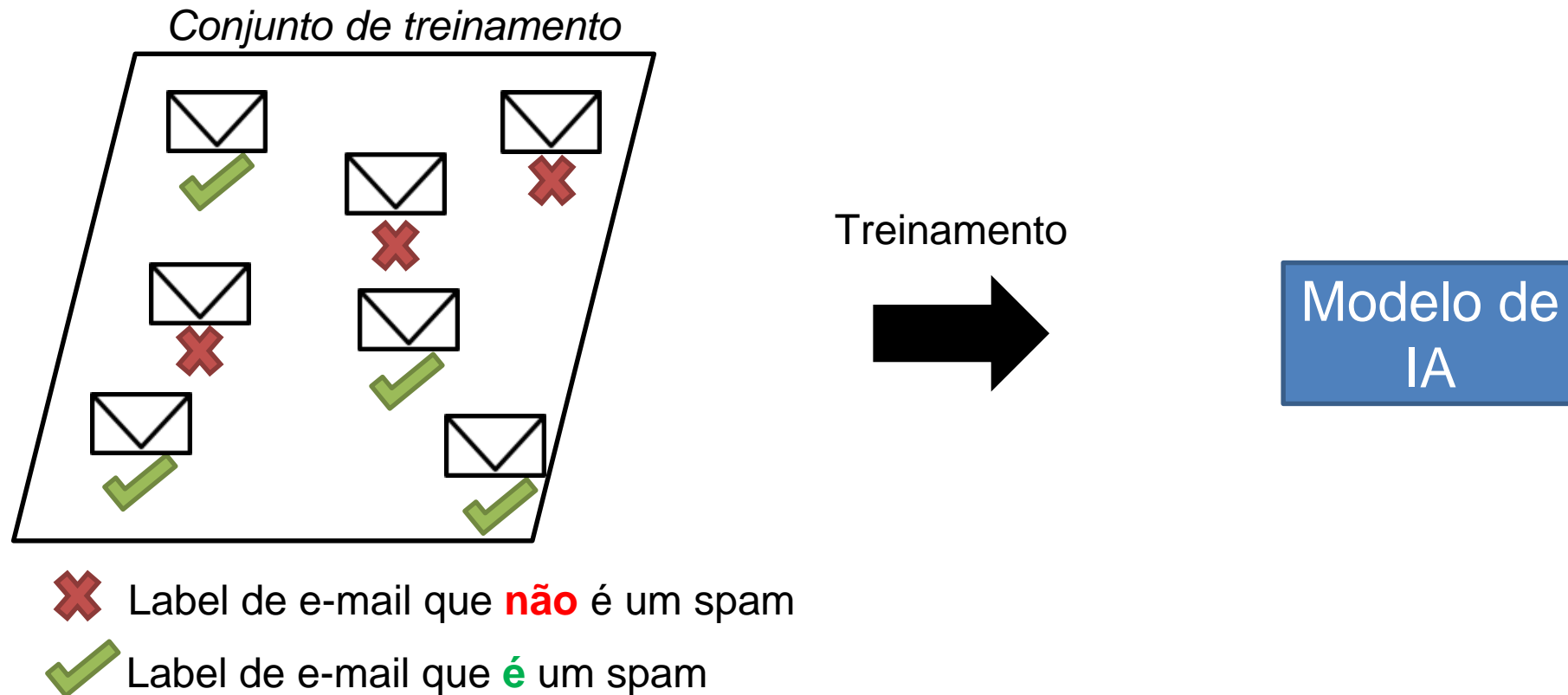
- Podemos ter um aprendizado do tipo **Supervisionado**



- Nesse tipo de aprendizado os dados do nosso problema possuem rótulos, ou em *labels*.
- Os rótulos são as saídas desejadas para o nosso problema
- Essas saídas desejadas e suas respectivas que irão auxiliar nosso algoritmo a aprender

Introdução à IA

- Podemos ter um aprendizado do tipo **Supervisionado**



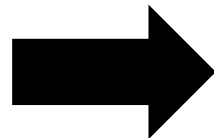
Introdução à IA

- Com o modelo de IA treinado podemos usá-lo com novos dados para obtermos

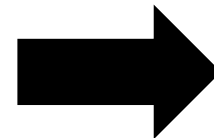
Novo e-mail



?

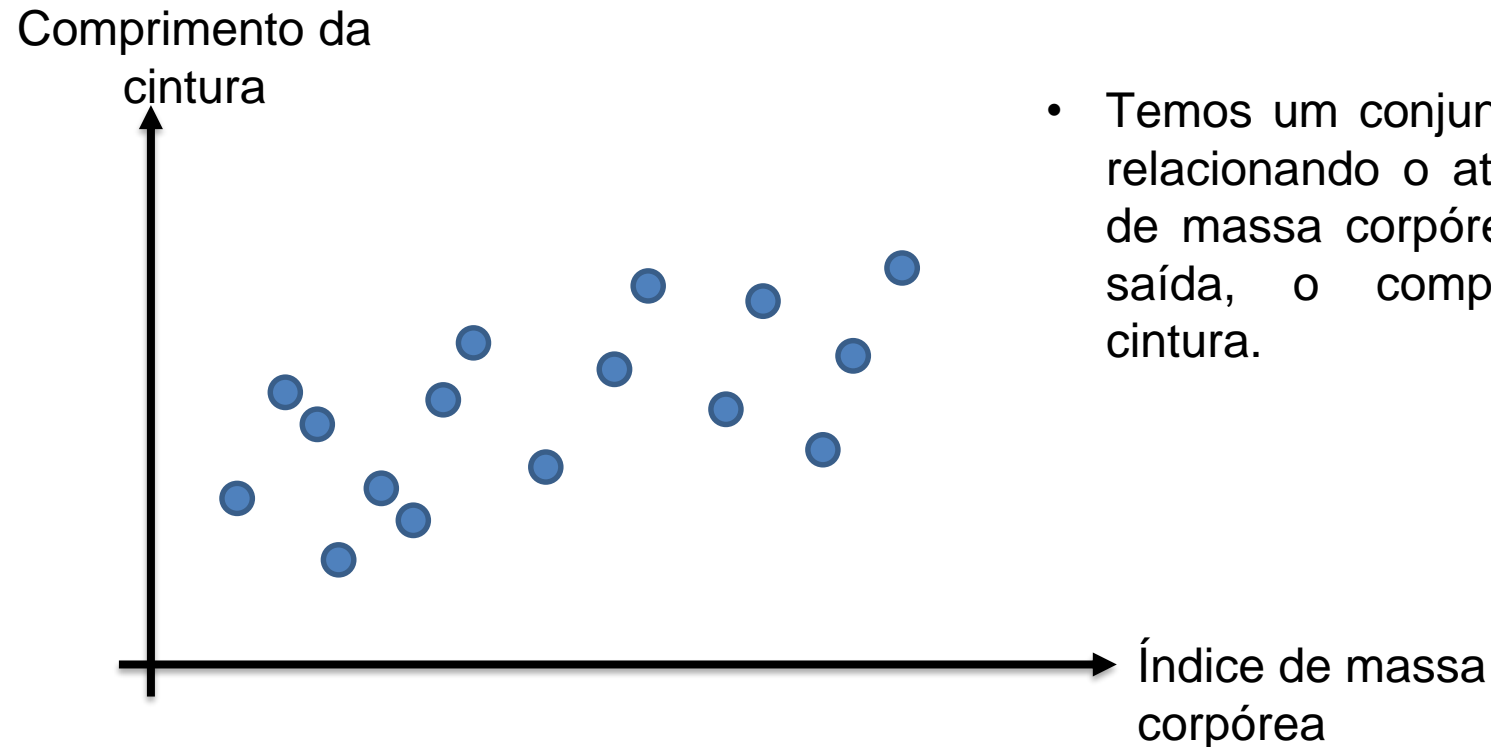


Modelo de
IA



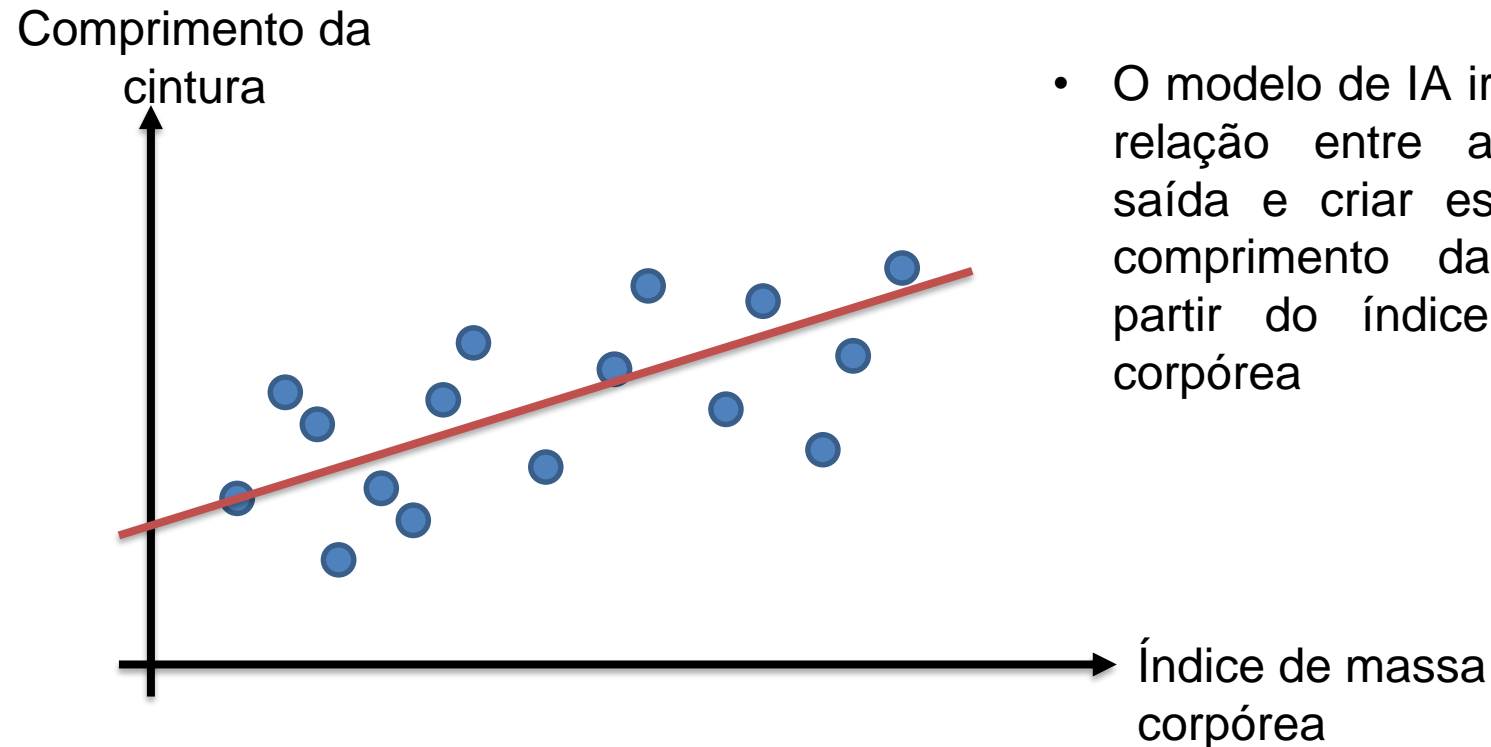
- Modelo aprendeu a relação entre as entradas e as saídas que foram apresentadas a ele na fase de treinamento

- Outra forma de trabalharmos com o aprendizado supervisionado é termos saídas com valores numéricos e não classes:



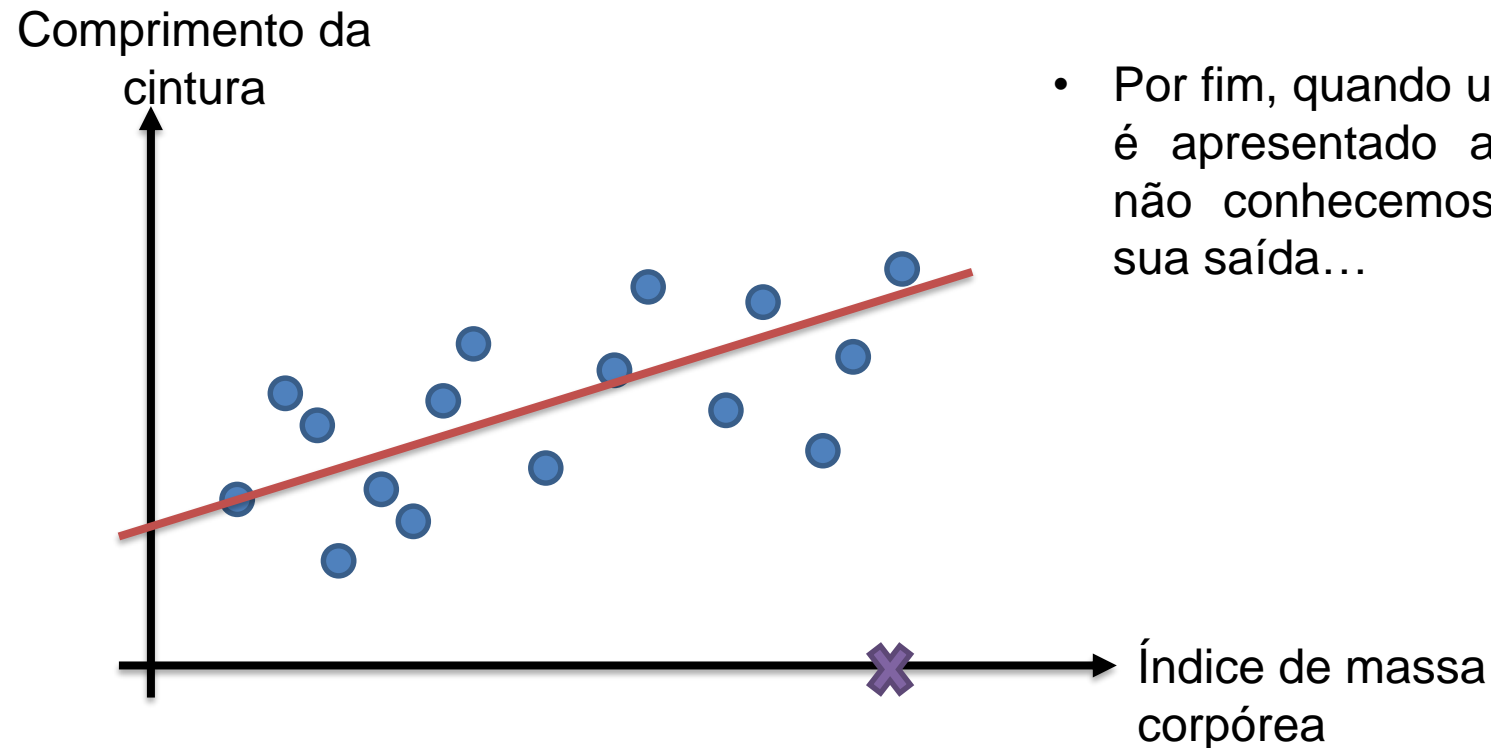
- Temos um conjunto de dados relacionando o atributo índice de massa corpórea com uma saída, o comprimento da cintura.

- Outra forma de trabalharmos com o aprendizado supervisionado é termos saídas com valores numéricos e não classes:



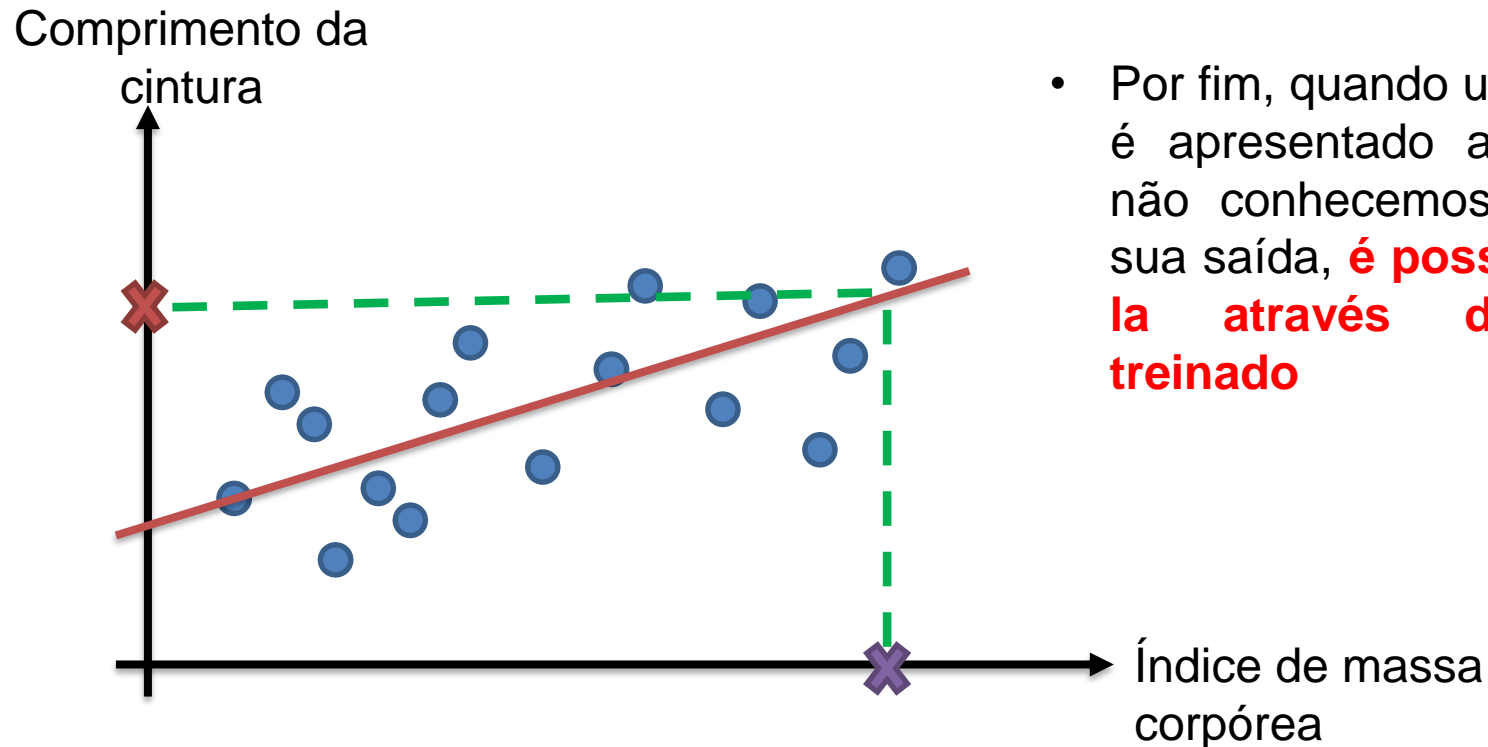
- O modelo de IA irá aprender a relação entre a entrada e saída e criar estimativas do comprimento da cintura, a partir do índice de massa corpórea

- Outra forma de trabalharmos com o aprendizado supervisionado é termos saídas com valores numéricos e não classes:



- Por fim, quando um novo dado é apresentado ao modelo e não conhecemos o valor de sua saída...

- Outra forma de trabalharmos com o aprendizado supervisionado é termos saídas com valores numéricos e não classes:

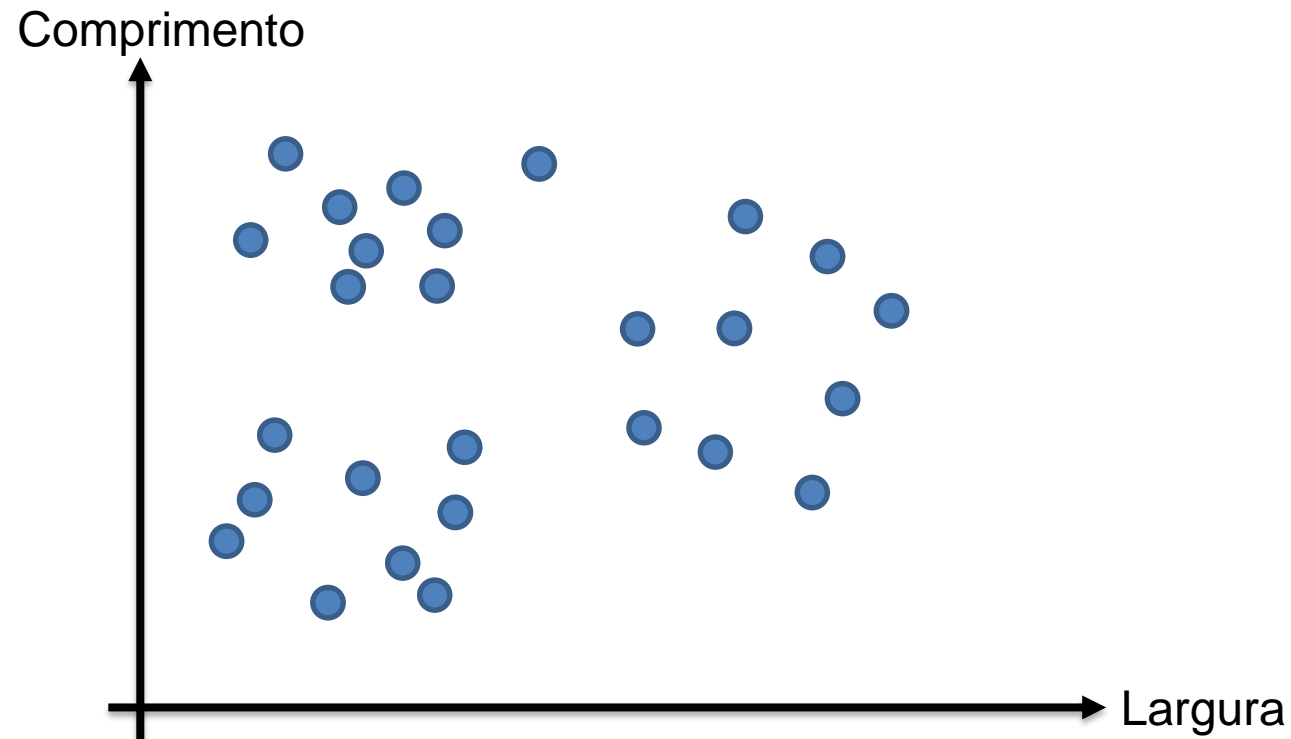


- Por fim, quando um novo dado é apresentado ao modelo e não conhecemos o valor de sua saída, **é possível estimá-la através do modelo treinado**

- Também temos o aprendizado do tipo **não Supervisionado**
- Nesse cenário buscamos encontrar padrões de agrupamento dos dados, uma vez que não temos os rótulos no nosso conjunto de dados

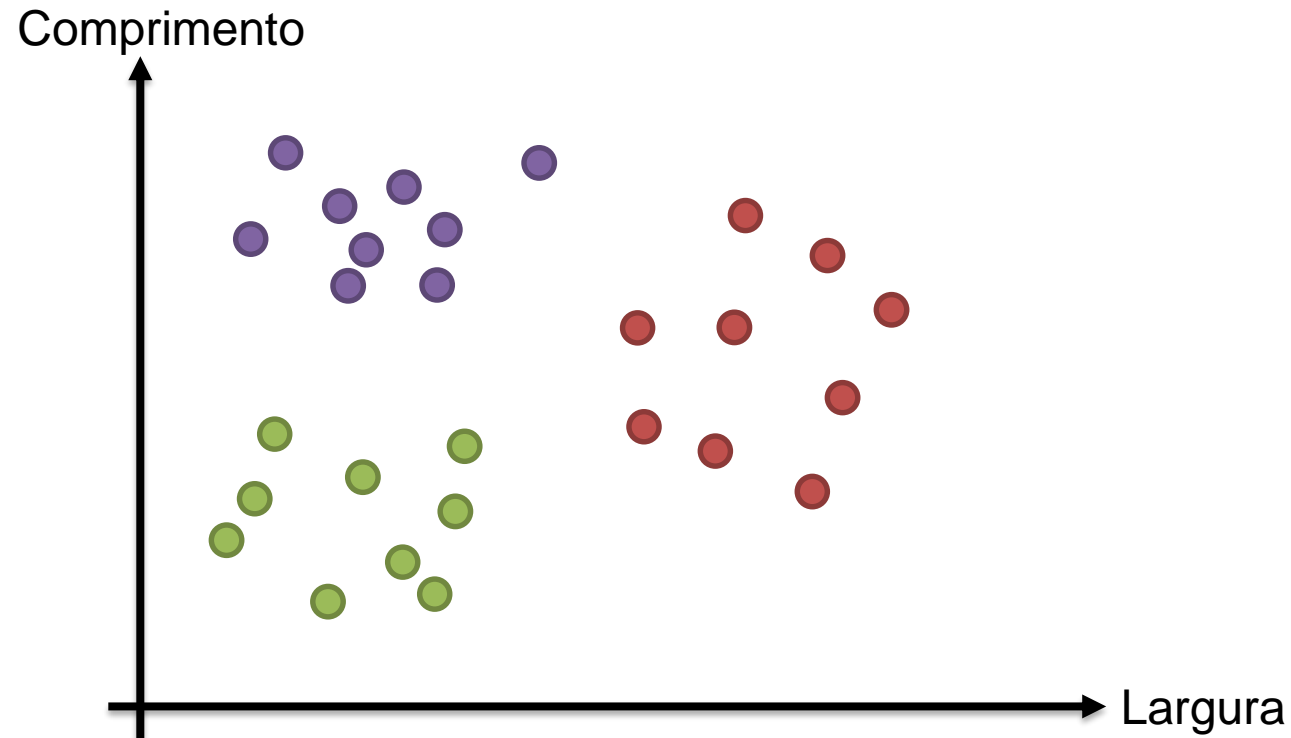
Introdução à IA

- Supondo que temos os dados a seguir de um conjunto de diferentes peixes. Não sabemos quais peixes foram medidos, apenas que são peixes.



Introdução à IA

- Entretanto, podemos inferir que temos três tipos de peixes de acordo com os agrupamentos a seguir. Isso é o aprendizado não supervisionado



Representação do Conhecimento

Representação do Conhecimento

- *Dados brutos* são uma coleção de entidades na sua forma bruta



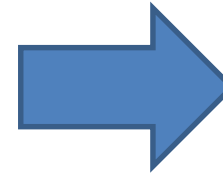
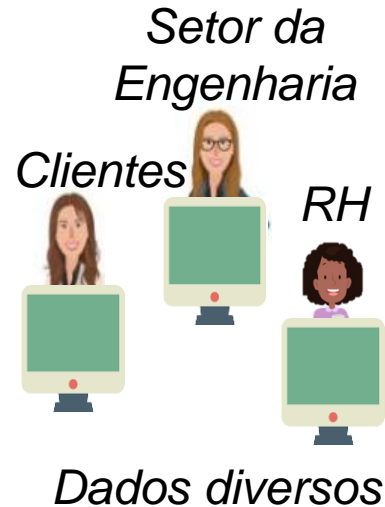
*Leitura de
instrumentos
musicais*



*Processos
Industriais*



*Imagens
armazenadas*



Dados brutos

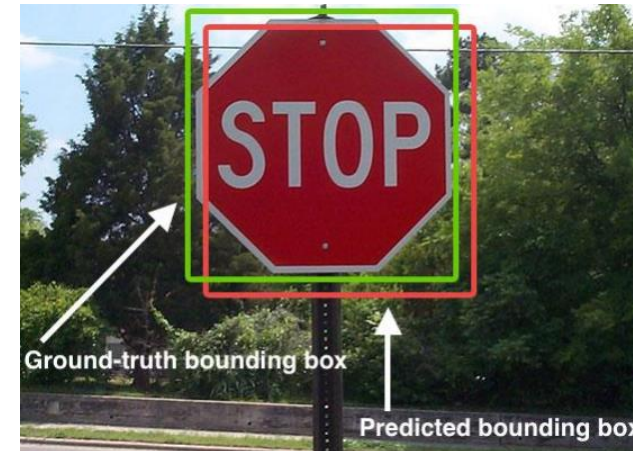
Representação da Informação

- Informação pode estar em diferentes mídias!

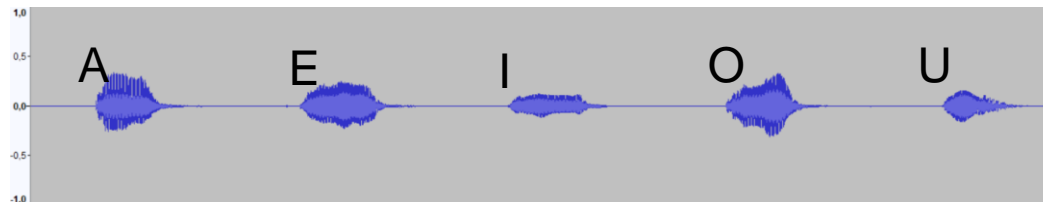
TABELA

Entrada	x_1	x_2	...	x_n	y	\hat{y}
1	70.52	30	...	0.584	90	100
2	60.96	27	...	1.254	81	90
...
k	97.48	35	...	0.758	122	120

IMAGEM



ÁUDIO



TEXTO

Esse campo de pesquisa ganhou muita notoriedade em 1986, quando David E. Rumelhart e James L. McClelland publicaram um livro que apresentou um modelo matemático computacional capaz de realizar um treinamento supervisionado dos neurônios artificiais. Esse algoritmo é chamado de **Backpropagation** e permite otimizações globais no modelo, sem restrições. Esse algoritmo também foi chamado de regra Delta generalizada, pois foi baseado na regra Delta, algoritmo de aprendizagem das redes Adalines.

Foi a partir desses trabalhos e da criação de diversos Journals e conferências que muitas instituições fundaram institutos de pesquisas e programas educacionais que estudam redes neurais artificiais e modelos de aprendizagem.

Nos próximos tópicos vamos aprender como a rede neural pode realizar predições através do algoritmo **Feedforward** e o aprendizado ou ajustes dos pesos, com o algoritmo **Backpropagation**. Animado? Vamos começar!

Representação do Conhecimento



- As técnicas de IA precisam de dados com o objetivo de manipulá-los para assim ser obtido algum conhecimento sobre eles.
 - Para que os algoritmos possam ser implementados precisamos ver as formas de representar o conhecimento

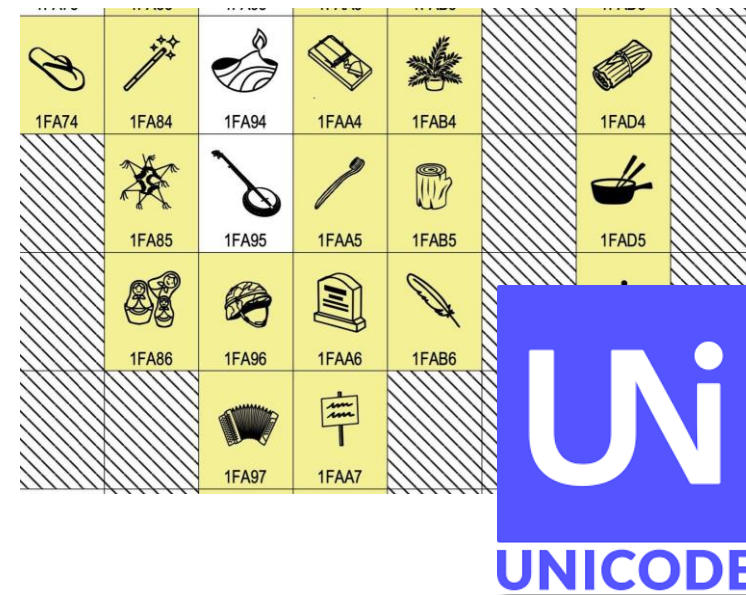
Representação do Conhecimento

- Esses dados brutos precisam de uma forma para serem representados para que posteriormente seja possível aplicar algum método de IA neles.



Representando Texto - Character Encoding

- Strings são objetos em linguagem de programação usados para trabalhar com caracteres;
- Os caracteres (e mais recentemente, emojis) são imagens mapeadas para um código hexadecimal (e binário);
- O mapeamento hexadecimal mais conhecido é o ASCII (American Standard Code II). Para contemplar outras línguas (além do alfabeto latino) e incorporar emojis, temos o Unicode;
- Em memória, o Unicode pode ser UTF-8, UTF-16, UTF-32;



Representando Texto - Embedding



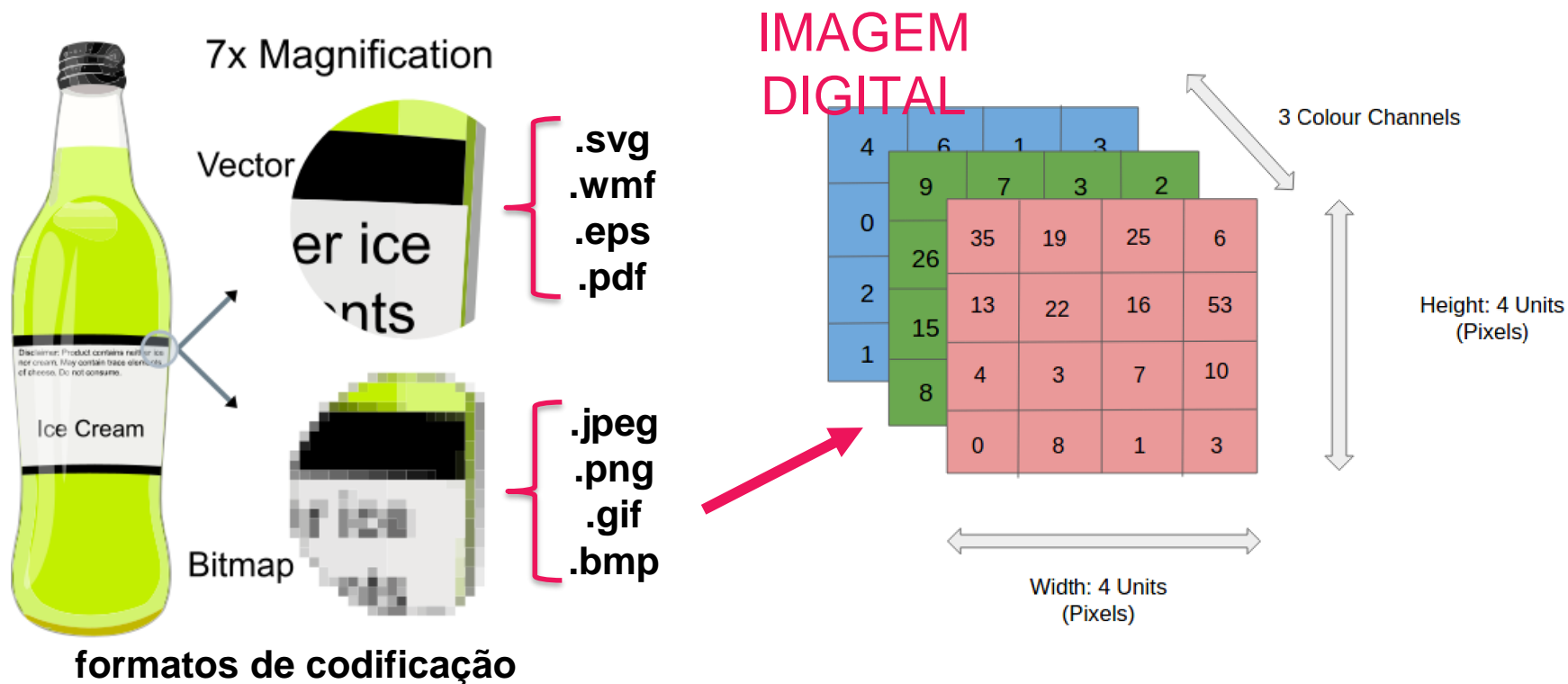
- Para algoritmos de **Inteligência Artificial**, as strings devem ser transformadas em outras representações numéricas;
- Uma técnica muito usada na área de **Processamento de Linguagem Natural** é transformar as string em vetores numéricos, uma técnica chamada de **Embedding**;
- Podemos ter Word Embedding quando representamos palavras por um vetor ou ainda Sentence Embedding quando representamos sentenças por vetores numéricos;
- Existem várias formas de fazer isso, entre elas:

Frases: Bag of Words (BOW) ou o TF-IDF

	cada	um	vou	eu	amanhã	cinema	em	hoje	e	a	ao	em
texto_vetor:	0	0	1	1	0	1	0	1	0	0	1	0

Representando Imagens

- Imagens digitais podem ter vários **formatos de codificação**;
- Imagens em formato RGB são bitmaps de 3 matrizes sobrepostas, onde cada elemento da matriz representa a intensidade daquele canal de cor naquela posição da imagem;
- **Resolução**: quantidade de pixels na altura e na largura;
- **Color depth**: quantidade de bits usados para cada número da matriz;
- Pixels próximos tendem a estar correlacionadas; já pixels distantes, não!



Representando Imagens

IMAGEM DIGITAL

Color depth: como armazenamos informação digital? Quantidade usada na memória física e dinâmica? Qualidade da representação?

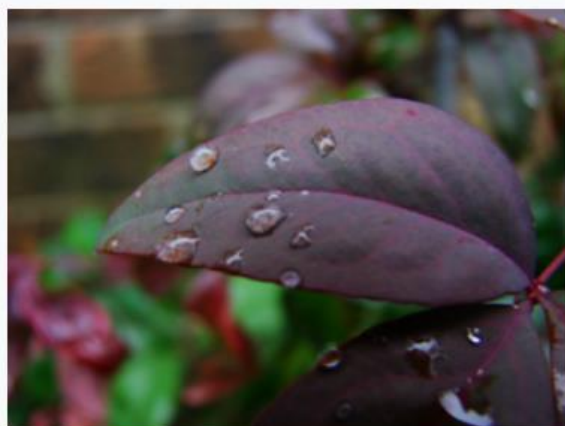


2 bit.png
4 colors
6 KB (-94%)

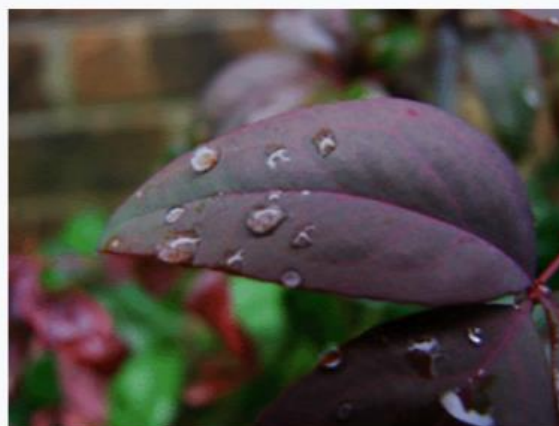


1 bit.png
2 colors
4 KB (-96%)

[2]



24 bit.png
16,777,216 colors
98 KB

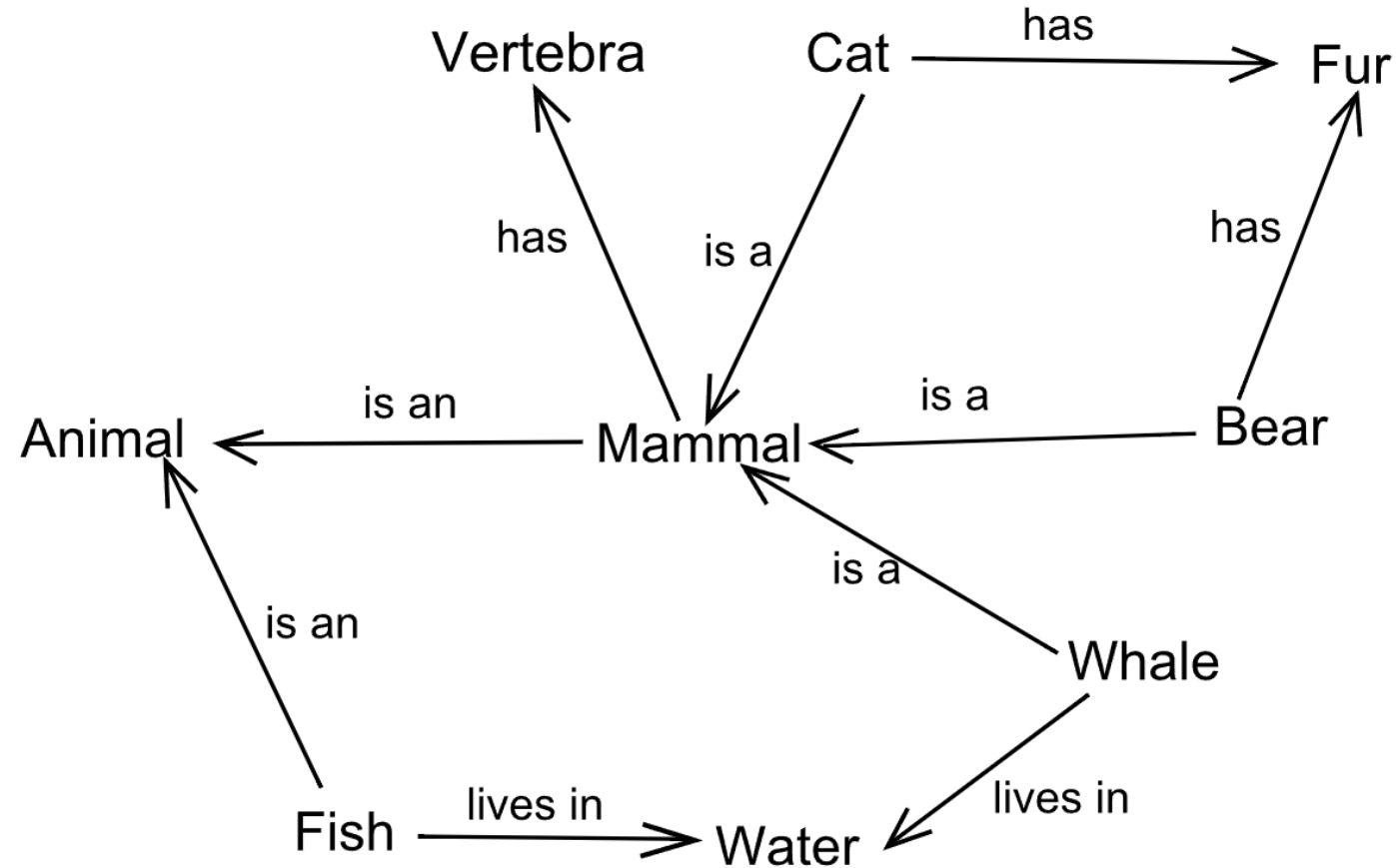


8 bit.png
256 colors
37 KB (-62%)



4 bit.png
16 colors
13 KB (-87%)

Representação por grafos



Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=1353062>

Representação de forma tabular



- Uma representação possível para os dados é através de organizar as suas informações em forma tabular:

model	engine_power	transmission	age_in_days	km	previous_owners	lat	lon	price
pop	69	manual	4474	56779	2	45.071079	7.46403	4490
lounge	69	manual	2708	160000	1	45.069679	7.70492	4500
lounge	69	automatic	3470	170000	2	45.514599	9.28434	4500
sport	69	manual	3288	132000	2	41.903221	12.49565	4700
sport	69	manual	3712	124490	2	45.532661	9.03892	4790

Representação de forma tabular

- Cada linha representa um **exemplo** dos nossos dados. Os exemplos são as diferentes entradas que foram coletadas de diversas entidades.

model	engine_power	transmission	age_in_days	km	previous_owners	lat	lon	price
pop	69	manual	1171	56779	2	45.071079	7.16403	1190
lounge	69	manual	2708	160000	1	45.069670	7.70492	4590
lounge	69	automatic	9470	170000	2	45.514599	9.20404	4590
sport	69	manual	3288	132000	2	41.903221	12.49585	4700
sport	69	manual	3712	124490	2	45.532661	9.03892	4790

- Nesse exemplo cada linha representa um carro distinto.

Representação de forma tabular

- Cada coluna representa um **atributo** dos nossos dados. Os atributos descrevem uma característica específica dos nossos exemplos

model	engine_power	transmission	age_in_days	km	previous_owners	lat	lon	price
pop	69	manual	1474	56779	2	45.071079	7.46403	4290
lounge	69	manual	2708	160000	1	45.069679	7.70492	4500
lounge	69	automatic	3470	170000	2	45.514599	9.28434	4500
sport	69	manual	3288	132000	2	41.903221	12.49565	4700
sport	69	manual	3712	124490	2	45.532661	9.03892	4790

- Nesse exemplo cada coluna representa características diferentes de cada carro.

Referências Bibliográficas



- BURKOV, Andriy. **Machine Learning Engineering**: Draft, 20XX.
- GÉRON, Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O'reilly Media, 2017