

# Aprendizado de Máquina

## Uma visão geral

Luiz Eduardo S. Oliveira

Universidade Federal do Paraná  
Departamento de Informática  
<http://web.inf.ufpr.br/luizoliveira>

# The world's most valuable resource is no longer oil, but data

*The data economy demands a new approach to antitrust rules*



CADE METZ BUSINESS 03.08.16 07:00 AM

# GOOGLE'S AI IS ABOUT TO BATTLE A GO CHAMPION—BUT THIS IS NO GAME

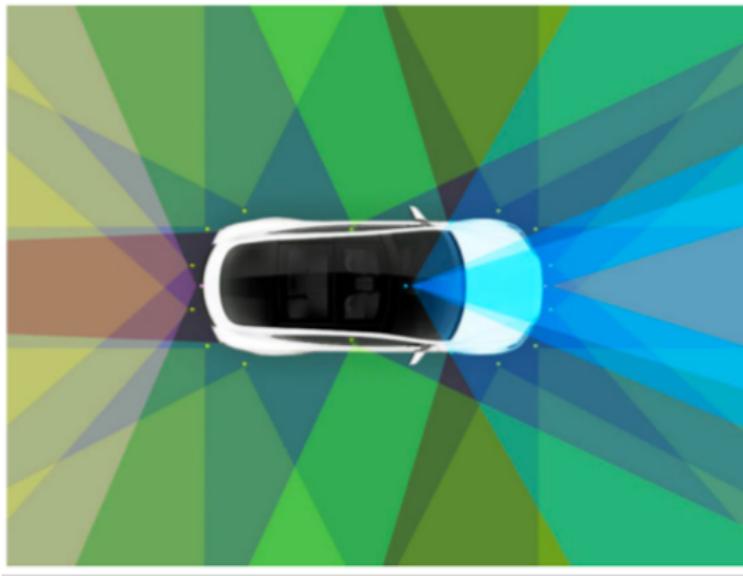


## Google's Go-playing AI still undefeated with victory over world number one

AlphaGo has won its second game against China's Ke Jie, sealing the three-game match in its favour



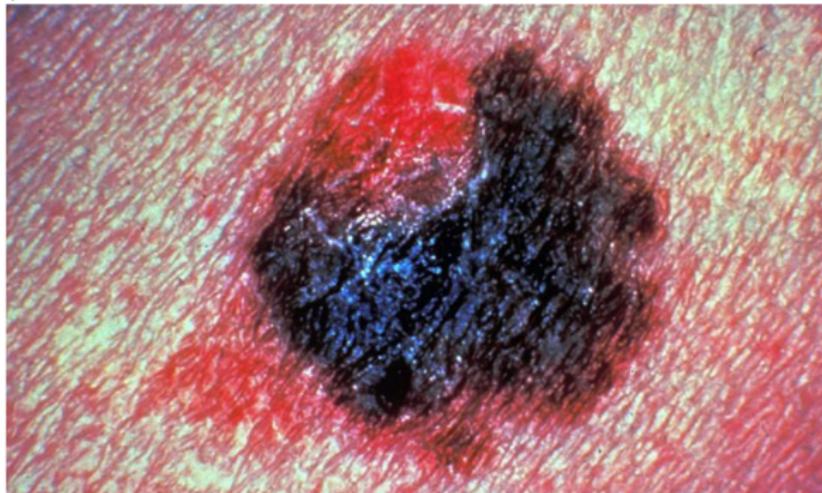
# TESLA FINALLY MAKES ITS NEW AUTOPilot AS GOOD AS THE OLD ONE





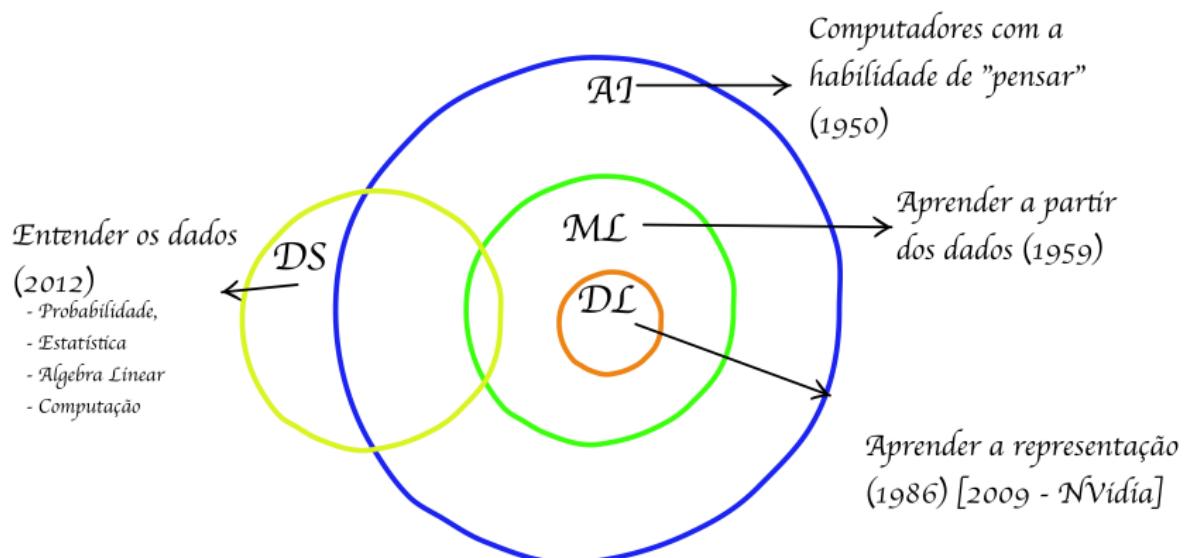
# This AI can spot skin cancer as accurately as a doctor

The artificial intelligence was trained on an image database of 129,000 images and performed as well as trained medical professionals



# Qual é a diferença?

- IA (Artificial Intelligence)
- ML (Machine Learning)
- DL (Deep Learning)
- DS (Data Science)



# Porque agora??

- Inteligência Artificial existe desde os anos 50.
- Disponibilidade de dados
- Poder computacional.
- Ferramentas.

# GPU



“To create the image and speech recognition algorithms designed by AutoML, Google reportedly let a cluster of **800** GPUs iterate and crunch numbers for weeks.”

[http://www.tomshardware.com/news/google-automl-artificial-intelligence-ai\\_34533.htm](http://www.tomshardware.com/news/google-automl-artificial-intelligence-ai_34533.htm)

# Machine Learning Frameworks



# Agenda

- O que é Aprendizagem de Máquina?
- Porque é importante?
- Diferentes abordagens de aprendizagem de máquina.
- Principais desafios.

# O que é Aprendizagem de Máquina?

- Campo de estudo que dá aos computadores a habilidade de aprender sem uma programação explícita [Arthur Samuel, 1959]

# O que é Aprendizagem de Máquina?

- Campo de estudo que dá aos computadores a habilidade de aprender sem uma programação explícita [Arthur Samuel, 1959]
- Pode-se dizer que um programa de computador aprende uma experiência E a respeito de uma tarefa T e alguma medida de desempenho P, se seu desempenho em T, medido por P, melhora com a experiência E. [Tom Mitchel, 1997]

# O que é Aprendizagem de Máquina?

- Campo de estudo que dá aos computadores a habilidade de aprender sem uma programação explícita [Arthur Samuel, 1959]
- Pode-se dizer que um programa de computador aprende uma experiência E a respeito de uma tarefa T e alguma medida de desempenho P, se seu desempenho em T, medido por P, melhora com a experiência E. [Tom Mitchel, 1997]
- É a ciência (e arte) de programar computadores de tal forma que eles aprendam a partir de dados [Aurélien Géron, 2017].

# Timeline

Credit: <https://i2.wp.com/>

## A.I. TIMELINE

**1950**

**TURING TEST**

Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence

**1955**

**A.I. BORN**

Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"

**1961**

**UNIMATE**

First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line



**1964**

**ELIZA**

Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans



**1966**

**SHAKY**

The 'first electronic person' from Stanford, Shaky is a general-purpose mobile robot that reasons about its own actions



**A.I.**

## WINTER

Many false starts and dead-ends leave A.I. out in the cold

**1997**

**DEEP BLUE**

Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov



**1998**

**KISMET**

Cynthia Breazeal at MIT introduces Kismet, an emotionally intelligent robot insofar as it detects and responds to people's feelings



**1999**

**AIBO**

Sony launches first consumer robot pet dog AIBO (AI robot) with skills and personality that develop over time



**2002**

**ROOMBA**

First mass produced autonomous robotic vacuum cleaner from iRobot learns to navigate and clean homes



**2011**

**SIRI**

Apple integrates Siri, an intelligent virtual assistant with a voice interface, into the iPhone 4S



**2011**

**WATSON**

IBM's question answering computer Watson wins first place on popular \$1M prize television quiz show Jeopardy



**2014**

**EUGENE**

Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human



**2014**

**ALEXA**

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks



**2016**

**TAY**

Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments



**2017**

**ALPHAGO**

Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number ( $2^{170}$ ) of possible positions

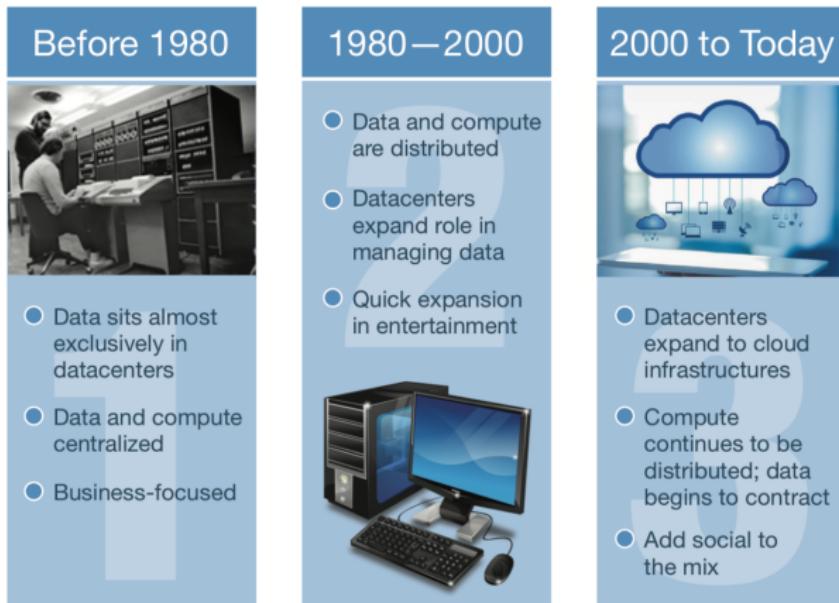


Porque é importante?



# Porque é importante?

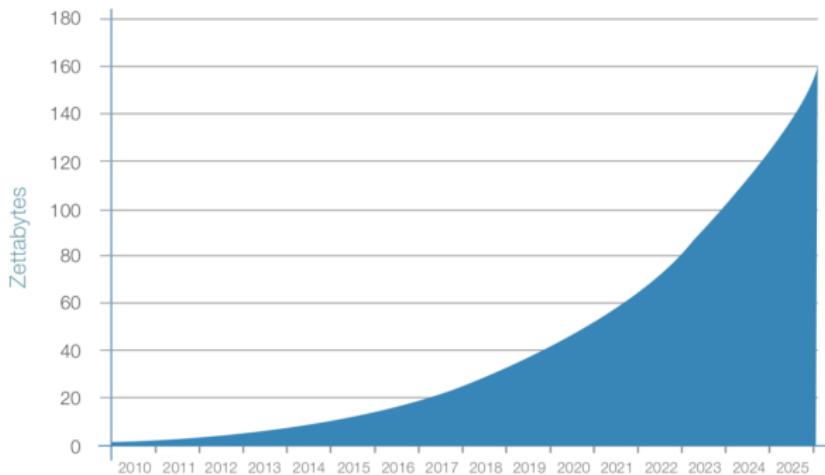
- Dados disponíveis em grandes escala



Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

# Porque é importante?

Dados criados



Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

- Impossível para o ser humano lidar com todo esse volume de dados.
- Imagine por exemplo a quantidade de recursos necessários para olhar e rotular cada imagem do Flickr ou Google Photos.
  - ▶ Aplicação real: Perícia em computadores e telefones.

# Desafios

- O aspecto chave da inteligência é a versatilidade, ou seja, a capacidade de fazer diversas coisas.
- Os sistemas de IA atualmente são muito bons e superam os seres humanos em uma atividade
  - ▶ Go
  - ▶ Jeopardy
  - ▶ Diagnóstico médico (detecção de cancer de pele)
  - ▶ etc...
- Entretanto, o sistema campeão de Jeopardy não consegue manter uma conversa.



# Quando é necessário utilizar aprendizagem de máquina?

# Quando é necessário utilizar aprendizagem de máquina?

Sempre que não for possível escrever um algoritmo determinístico para resolver o problema. Nesses casos o conhecimento deve ser extraído a partir de exemplos.

## Amazon Reviews:

- Negativos:

- ▶ The Worst!: A complete waste of time. Typographical errors, poor grammar, and a totally pathetic plot add up to absolutely nothing. I'm embarrassed for this author and very disappointed I actually paid for this book.
- ▶ Awful beyond belief!: I feel I have to write to keep others from wasting their money. This book seems to have been written by a 7th grader with poor grammatical skills for her age!
- ▶ Oh dear: I was excited to find a book ostensibly about Muslim feminism, but this volume did not live up to the expectations

- Positivos:

- ▶ Even Mommy has fun with this one!: My four year old daughter loves everything Barbie and loves the Rapunzel movie. This game is tons of fun, even for a 42 year old.
- ▶ Rapunzel: This is such a great game both my 3 year old son and 7 year old daughter love it. I like to play to if they would let me!

# Exemplos de Aplicações



# Exemplos

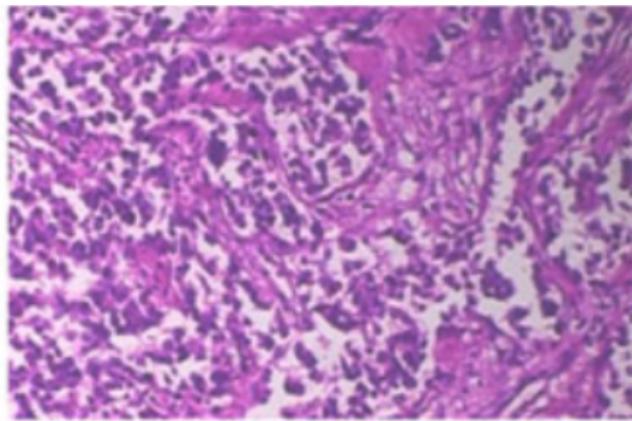
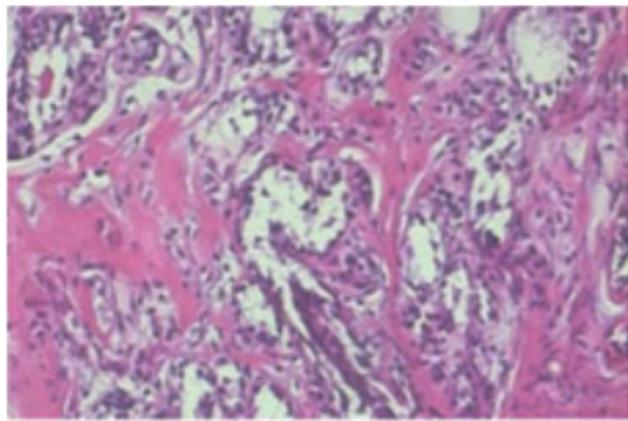
## Detecção de objetos



# Exemplos

Imagens médicas.

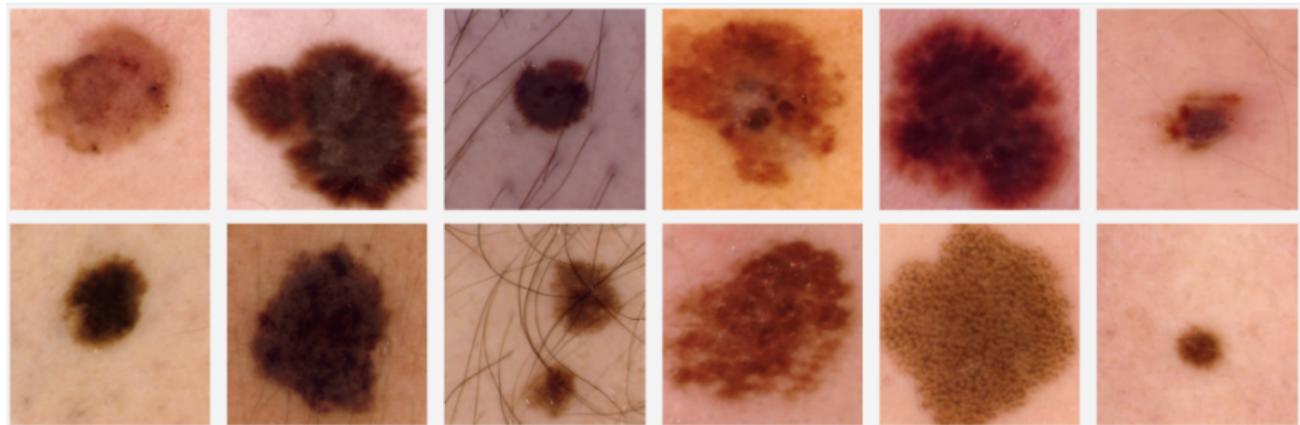
- Especialistas humanos altamente treinados são capazes de classificar essas imagens.
- Transferir conhecimento.



# Exemplos

Imagens médicas.

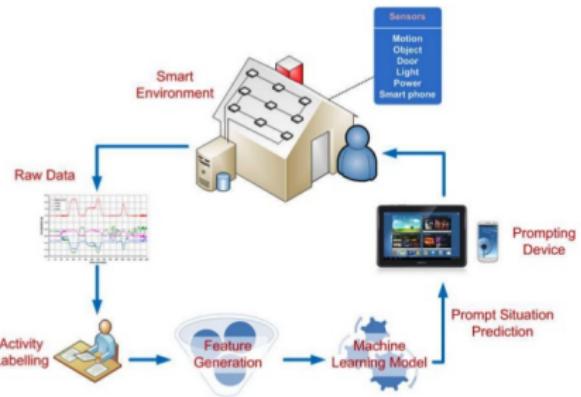
- Melanomas (linha de cima) e lesões de pele benignas (linha de baixo)



# Exemplos

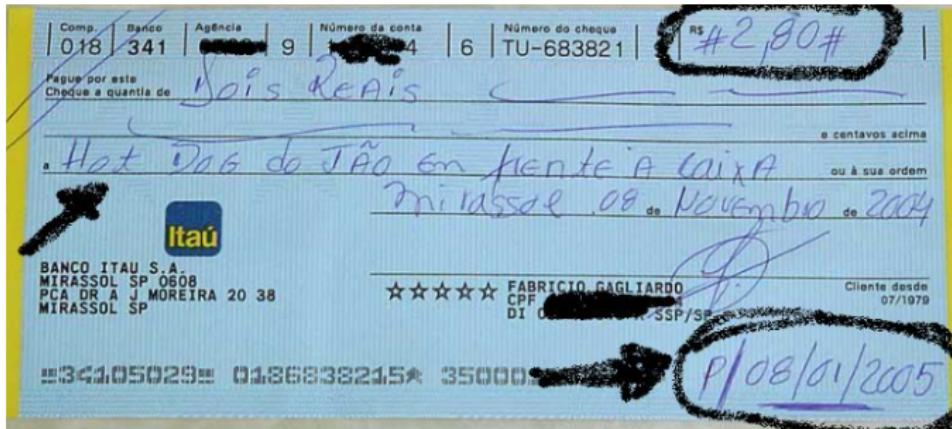
## Detecção de Anomalia

- Uso de cartão de crédito.
- Redes de computadores.
- Sistemas de monitoração.
- Smart Homes, etc

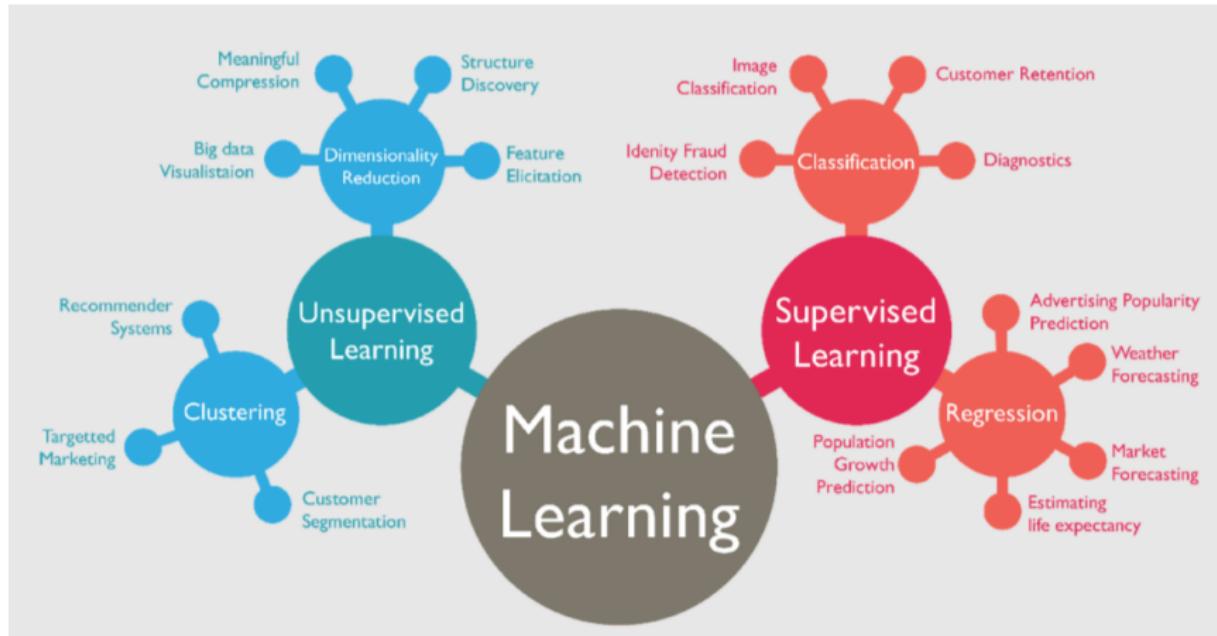


# Exemplos

## Reconhecimento de Manuscritos

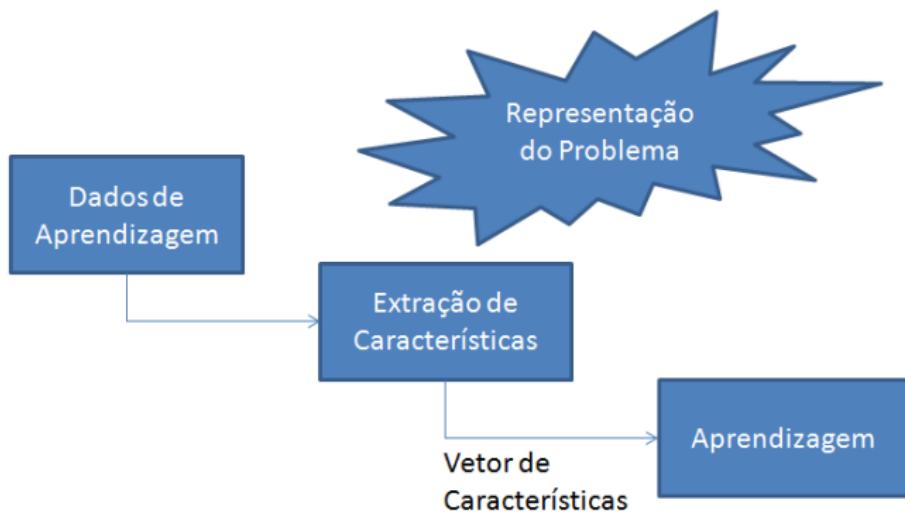


# Tipos de Aprendizagem



- Supervisionada: dados de treinamento estão rotulados.
- Não supervisionada: Sem rótulos; descobrir estruturas.

# Estrutura Básica - Aprendizagem Supervisionada



# Extração de Características (Representação)

## Objetivo

Caracterizar um objeto através de medidas, as quais são bastante similares para objetos da mesma classe, e bastante diferentes para objetos de outras classes.

- Características devem ser discriminantes e invariantes.
- A maioria dos modelos de aprendizagem exigem que as características sejam normalizadas.

# Extração de Características (Representação)

Qualquer medida que se possa extrair de um determinado objeto.

- Simbólicas
  - ▶ Ex: Cor
- Numéricas contínuas
  - ▶ Peso
- Binárias
  - ▶ Determinam a presença ou ausência de uma determinada característica.

Alguns algoritmos de aprendizagem de máquina são mais indicados para determinados tipos de características.

# Extração de Características

## Globais

- Extrair características de uma maneira holística.
- Maneira como os humanos reconhecem objetos
- Gestalt (Percepção)
- Em alguns casos a segmentação para extração de características locais não é fácil, como por exemplo, texturas.

## Locais

- Segmentar em partes menores para então extrair características.
- Identificar a estrutura do objeto a ser reconhecido e então extrair partes de interesse.

# Percepção

Processo de adquirir, interpretar, selecionar e organização informações sensoriais

## Gestalt

- Percebemos objetos como padrões bem organizados ao invés de componentes separados
- Baseado no conceito de “agrupamento” e nos seguintes princípios:
  - ▶ Emergência
  - ▶ Construtivismo
  - ▶ Invariância

## Emergência



## Emergência



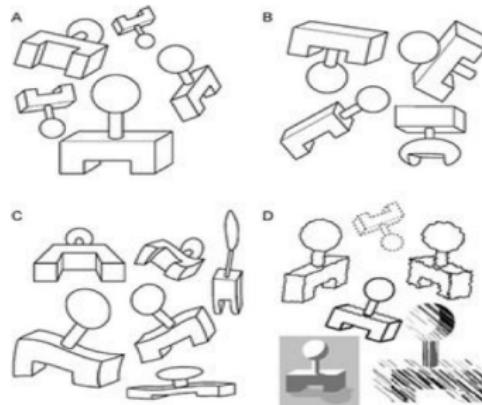
# Gestalt

## Construtivismo



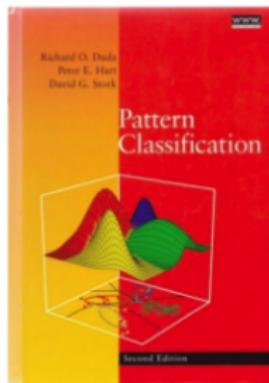
Conseguimos identificar características não presentes a partir da percepção de características presentes.

## Invariância



Objetos são reconhecidos independentemente de rotação, translação, escala e ruído.

# Aprendizagem Supervisionada - Classificação



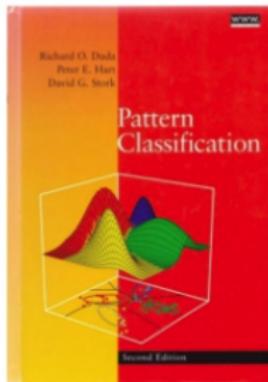
Robalo



Salmão

- Construir um sistema que classifique automaticamente dois tipos de peixe: Salmão e Robalo.
  - ▶ Precisamos de aprendizagem de máquina?

# Aprendizagem Supervisionada - Classificação



Robalo



Salmão

- Construir um sistema que classifique automaticamente dois tipos de peixe: Salmão e Robalo.
  - ▶ Precisamos de aprendizagem de máquina?
  - ▶ Sim! O que precisamos para construir o sistema?

# Aprendizagem Supervisionada - Classificação

## 1. Dados Rotulados



- Esse conjunto de dados rotulados é chamado de base de treinamento.

# Aprendizagem Supervisionada - Classificação

## 2. Representação

- O que diferencia um salmão de um robalo?
- Extrair atributos, os quais sejam similares aos objetos de uma mesma classe e diferentes para objetos das outras classes.
- Processo conhecido com **Extração de Características**



# Aprendizagem Supervisionada - Classificação

## 2. Representação

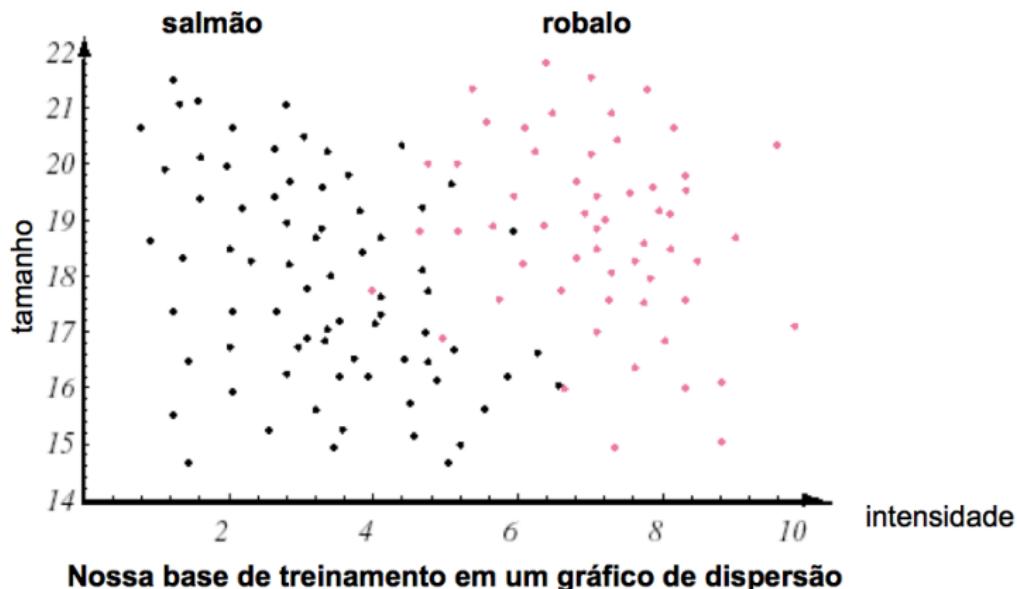
- O que diferencia um salmão de um robalo?
- Extrair atributos, os quais sejam similares aos objetos de uma mesma classe e diferentes para objetos das outras classes.
- Processo conhecido com **Extração de Características**



- Suponha que tamanho e cor (intensidade) sejam características discriminantes.

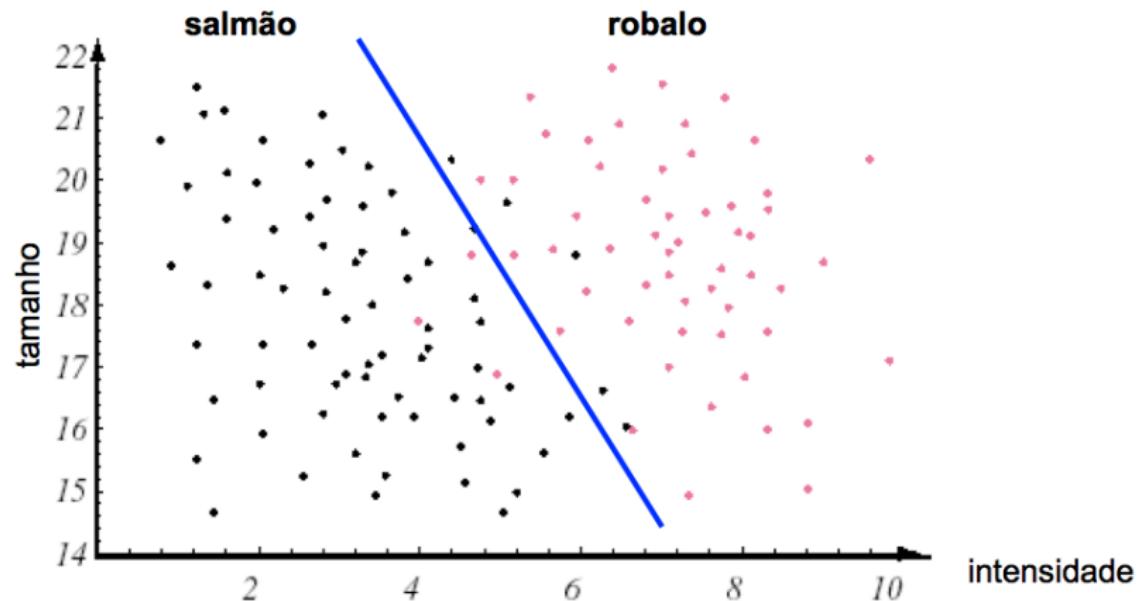
# Aprendizagem Supervisionada - Classificação

## 2. Representação



# Aprendizagem Supervisionada - Classificação

## 3. Fronteira de decisão



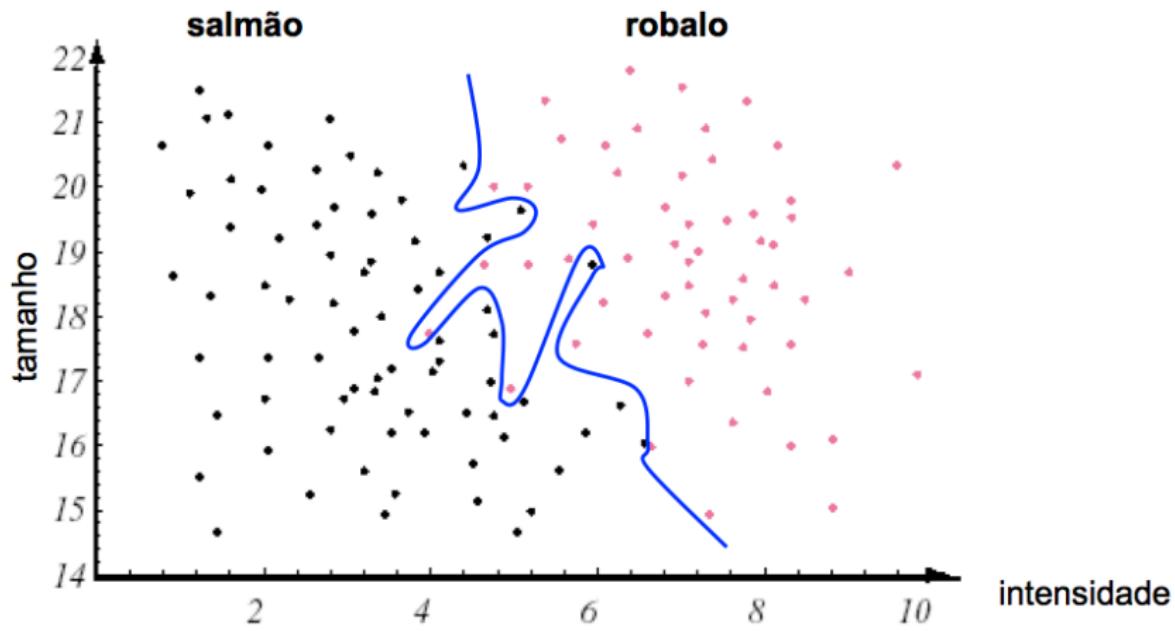
- Separação Linear:  $y = ax + b$
- Funções Discriminantes Lineares (Perceptron, SVM).

## 3. Fronteira de decisão

- Essa regra fornece uma boa fronteira de decisão, entretanto pode ser melhorada.
- Adicionar mais características, como por exemplo, dimensões do peixe, posição relativa dos olhos, etc..., poderiam ajudar.
- Quanto mais características, mais base de treinamento será necessária
- Maldição da dimensionalidade

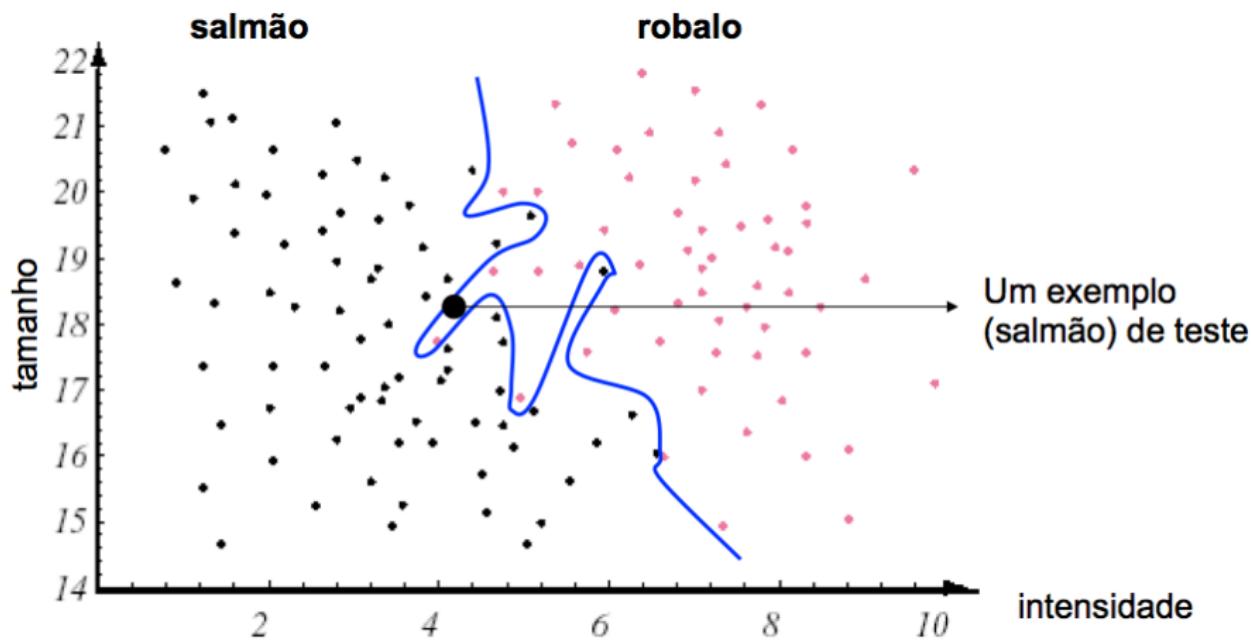
# Aprendizagem Supervisionada - Classificação

## 3. Fronteira de decisão



Aprendizagem Supervisionada - Classificação

### 3. Fronteira de decisão



# Aprendizagem Supervisionada - Classificação

## 3. Fronteira de decisão

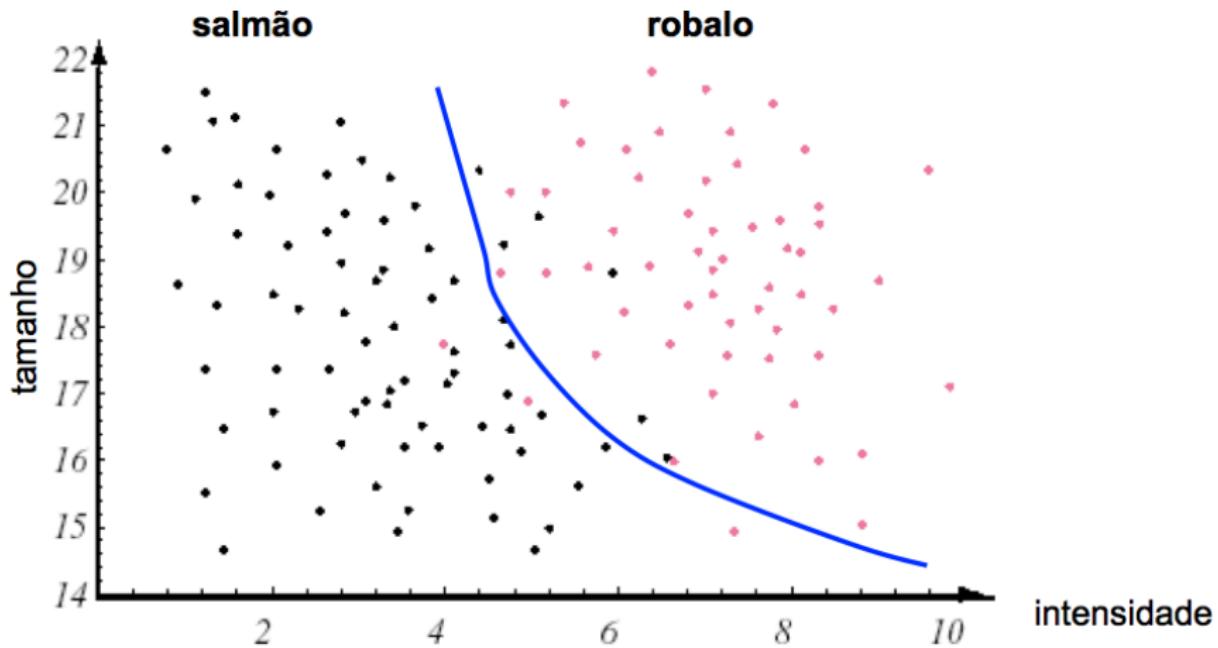
- Apesar de complexo, esse modelo é pior do que o modelo linear.
- A principal característica de um modelo deve ser a sua capacidade de generalizar.
- Modelos muito complexos geralmente não generalizam bem, pois decoram a base de treinamento (over-fitting).

## 3. Fronteira de decisão

- Por outro lado, modelos muito simples tendem a generalizar demais (under-fitting).
- Na aprendizagem de máquina, veremos técnicas para evitar over- e under-fitting, ou seja, construir fronteiras de decisão nem tão simples e nem tão complexas

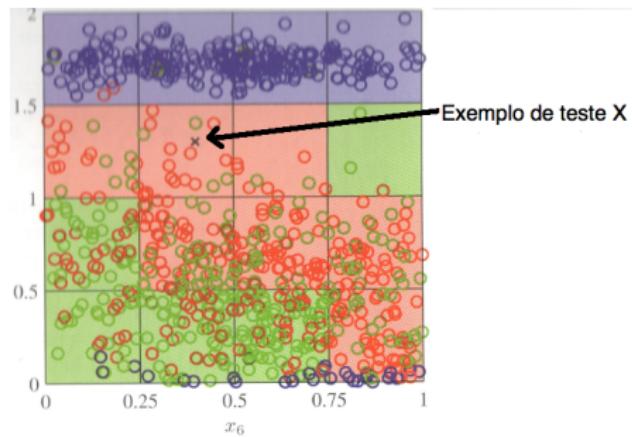
# Aprendizagem Supervisionada - Classificação

## 3. Fronteira de decisão



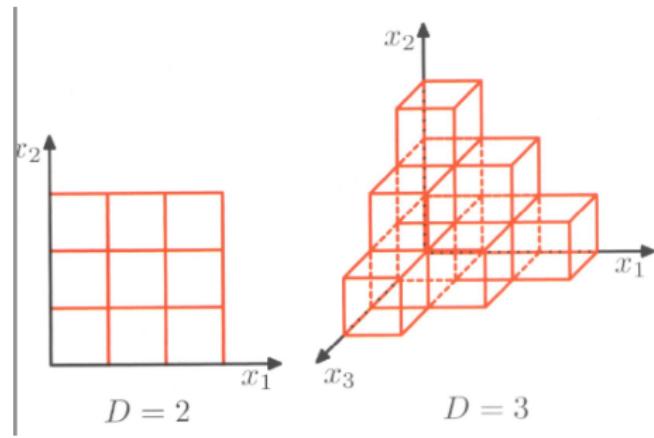
# Maldição da Dimensionalidade

- Em geral, na busca por melhores resultados na aprendizagem, criam-se vetores de características com mais informações.
- Para ilustrar o problema, vamos considerar o exemplo da Figura abaixo, com duas características  $x_6$  e  $x_7$  e três classes.
- Para classificar o padrão  $\mathbf{x}$ , dividimos o espaço em células de tamanho igual e atribuímos a  $\mathbf{x}$  a classe mais frequente dentro da célula.



# Maldição da Dimensionalidade

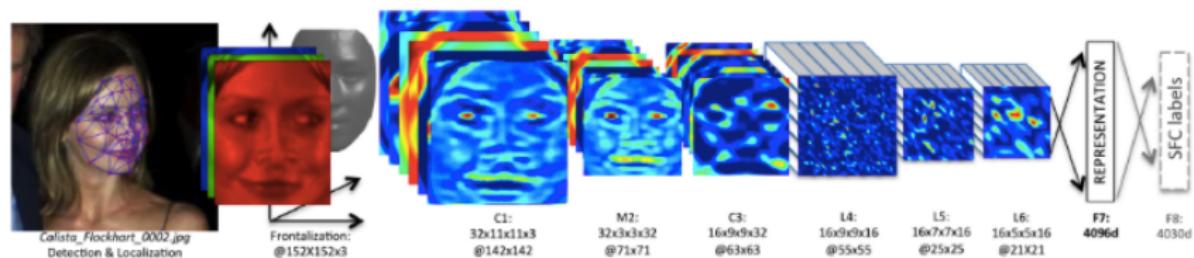
- Conforme adicionamos características ao vetor, o número de células cresce de forma exponencial.
- Em função disso, quanto mais características temos, mais base de dados temos que ter para preencher todas as células.
- Na maioria dos problemas, entretanto, a quantidade de dados disponível para a aprendizagem é limitada.



# Aprendizagem Supervisionada - Classificação

## Deep Learning (Representation Learning)

- Extrair a representação a partir dos dados.
- Qualquer algoritmos de aprendizagem de máquina depende da representação usada.
- Definição de características não é uma tarefa trivial.
- Porque não deixa-la então por conta do algoritmo de aprendizagem?

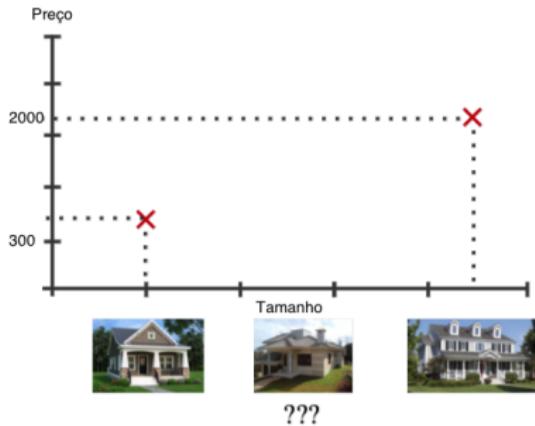


# Aprendizagem Supervisionada - Regressão



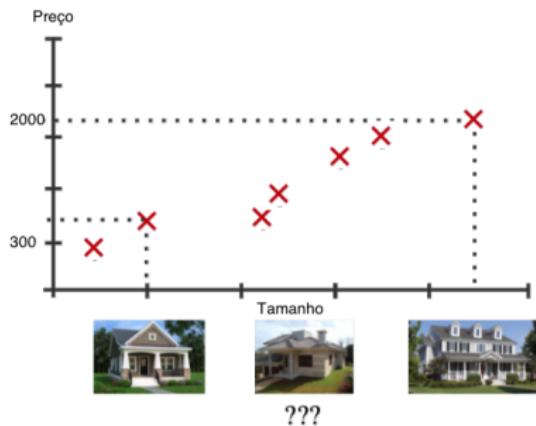
- Construir um sistema para estimar do preço de um imóvel.

# Aprendizagem Supervisionada - Regressão



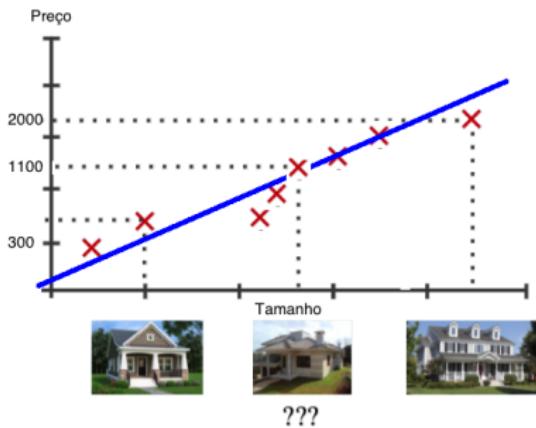
- Com base em informações de preço e tamanho de diversas casas, a tarefa consiste em estimar o preço de uma dada casa.
- Diferentemente da classificação, onde a saída do classificador é categórica, na regressão a saída é um valor real.

# Aprendizagem Supervisionada - Regressão



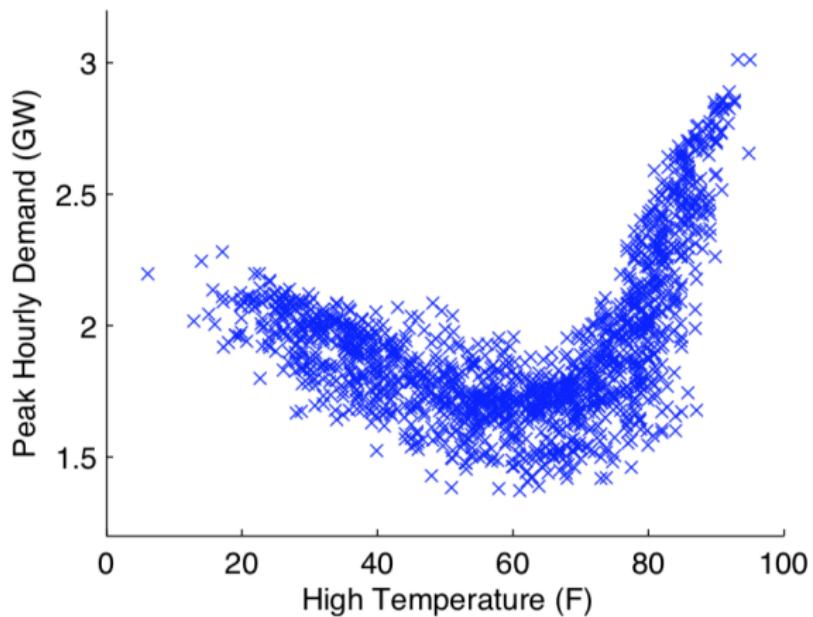
- Qual seria o preço estimado para a casa em questão?

# Aprendizagem Supervisionada - Regressão



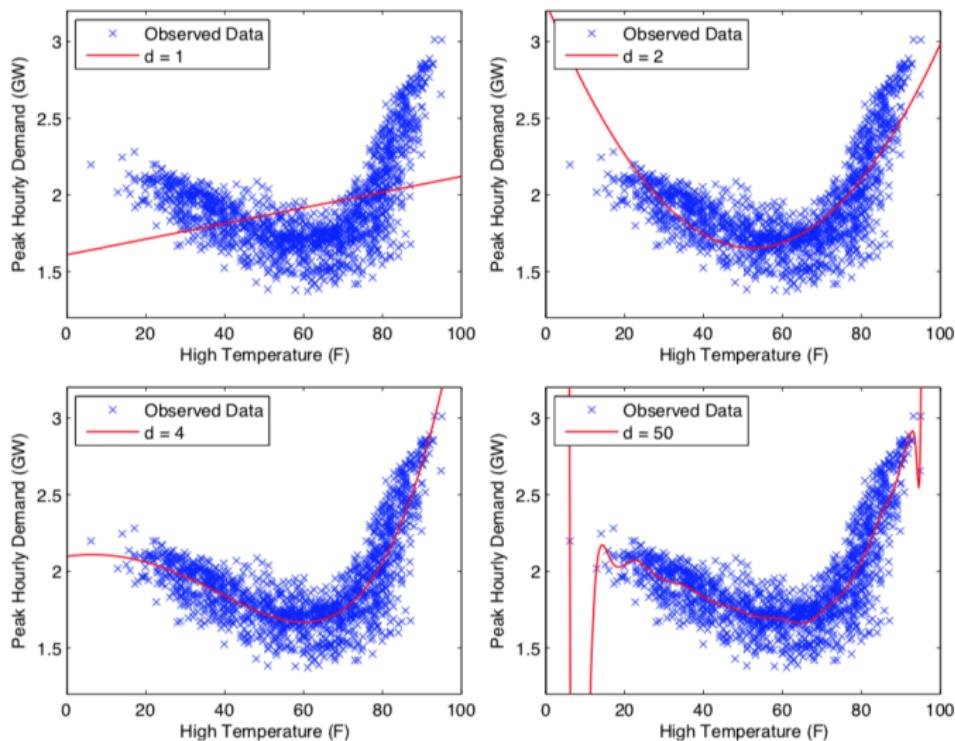
- Dado um conjunto de dados rotulados (tamanho da casa, preço) podemos encontrar um modelo para estimar o preço de outras casas.
- O modelo mais simples é a regressão Linear ( $y = ax + b$ )
- Cuidado para não extrapolar sua base de conhecimento.

# Aprendizagem Supervisionada - Regressão



- Em alguns casos um modelo linear não é a melhor solução.

# Aprendizagem Supervisionada - Regressão



- Under-fitting vs Over-fitting

# Aprendizagem não supervisionada

- Em problemas de aprendizagem supervisionada, contamos como a tupla  $[X, y]$ , em que o objetivo é classificar  $y$  usando o vetor de características  $X$ .
- Na aprendizagem não supervisionada, temos somente o vetor  $X$ .
- Nesse caso, o objetivo é descobrir alguma coisa a respeito dos dados
  - ▶ Por exemplo, como eles estão agrupados.
- Mais subjetiva que a aprendizagem supervisionada uma vez que não existe um objetivo simples como a classificação.

## Dados não rotulados

Obtenção de dados não rotulados não é custosa!

# Aprendizagem não supervisionada

Exemplos de aplicações:

- Segmentação de mercado (Marketing)
- Agrupamento de pacientes|clientes|documentos, etc...
- Sistemas de Recomendação

Frequently Bought Together



This item: Structure and Interpretation of Computer Programs - 2nd Edition (MIT Electrical Engineering and... by Harold Abelson Paperback \$50.50

The Pragmatic Programmer: From Journeyman to Master by Andrew Hunt Paperback \$32.59

Customers Who Bought This Item Also Bought

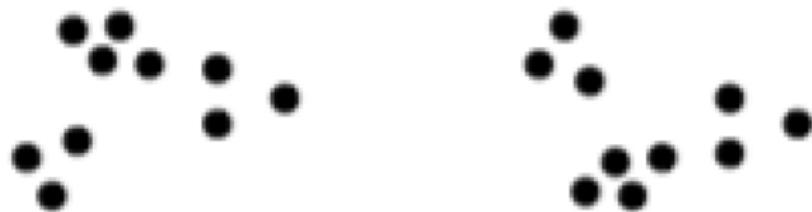
The screenshot shows a grid of recommended books for 'The Little Schemer'. The books listed are:

- The Little Schemer - 4th Edition by Daniel P. Friedman (4.5 stars, \$64 Prime)
- Instructor's Manual to Structure and Interpretation of Computer Programs... by Gerald Jay Sussman (4.5 stars, \$28.70 Prime)
- The Pragmatic Programmer by Andrew Hunt (4.5 stars, \$32.59 Prime)
- Introduction to Algorithms, 3rd Edition (MIT Press) by Thomas H. Cormen (4.5 stars, \$66.32 Prime)
- An Introduction to Functional Programming Through Lambda Calculus by Greg Michaelson (4.5 stars, \$20.70 Prime)
- Purely Functional Data Structures by Chris Okasaki (4.5 stars, \$40.74 Prime)
- Code: The Hidden Language of Computer Hardware and Software by Charles Petzold (4.5 stars, \$17.99 Prime)
- The Little Prover (MIT Press) by Daniel P. Friedman (4.5 stars, \$31.78 Prime)

Page 1 of 13

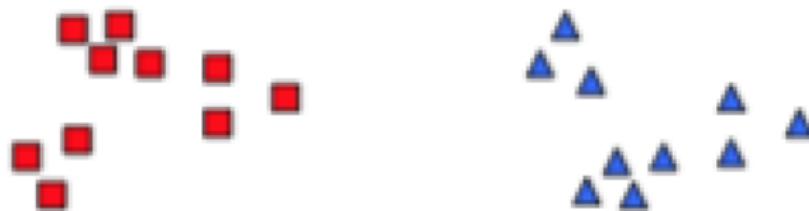
# Aprendizagem não supervisionada

Quantos grupos (clusters) existem nesses dados?



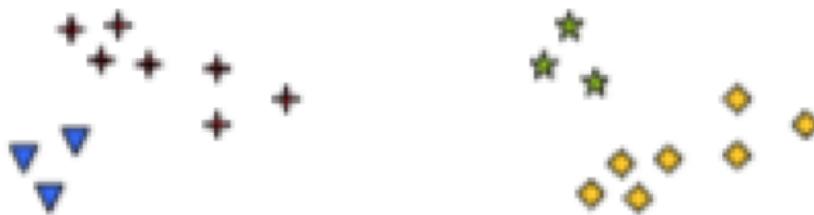
# Aprendizagem não supervisionada

Dois



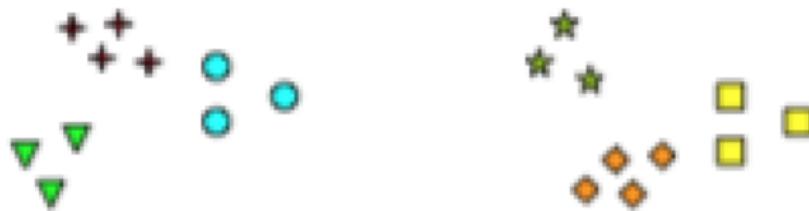
# Aprendizagem não supervisionada

Quatro



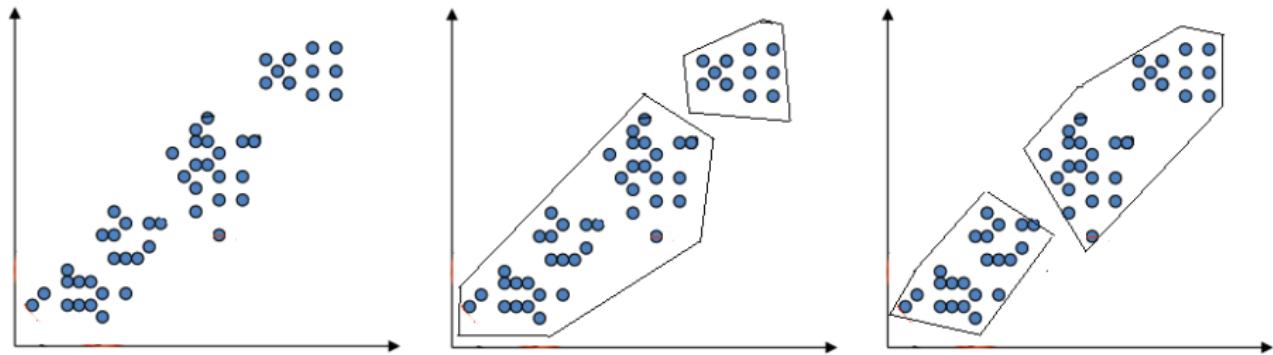
# Aprendizagem não supervisionada

Seis



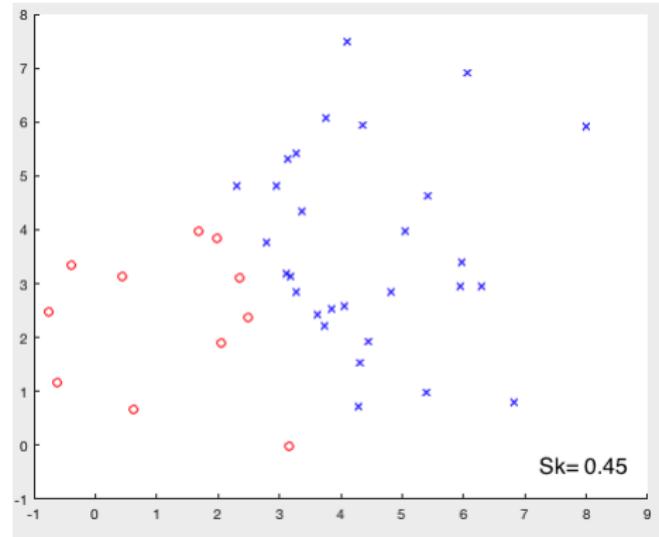
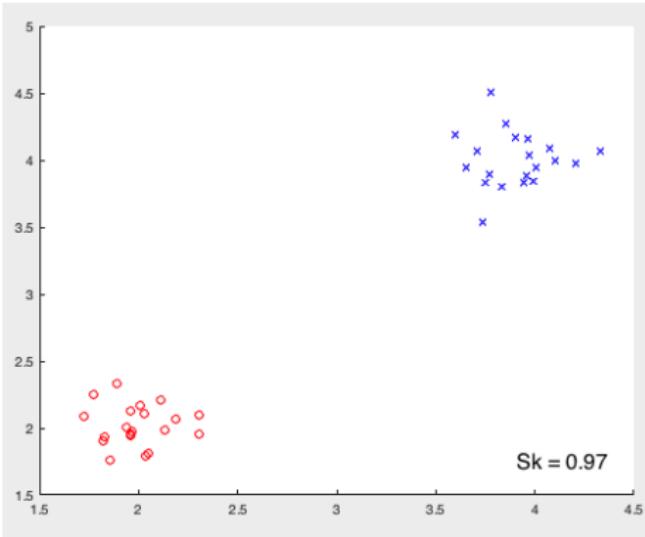
# Aprendizagem não supervisionada

O que é um bom clustering?



- Compactos e longe um dos outros.
- Índices de clusters

# Avaliação em Aprendizagem não Supervisionada



# Redução de Dimensionalidade

- Independentemente do algoritmo de aprendizagem de máquina escolhido, um dos principais aspectos na construção de um bom classificador é a utilização de características discriminantes.
- A adição de uma nova característica não significa necessariamente um bom classificador
- Depois de um certo ponto, adicionar novas características pode piorar o desempenho do classificador.
- Aspectos diretamente relacionados com a escolha das características:
  - ▶ Desempenho
  - ▶ Tamanho da base de dados.

Identificar e selecionar um subconjunto de características relevantes para um determinado problema, a partir de um conjunto inicial.

## Redução de Dimensionalidade

Cuidado para não ter uma visão parcial dos dados...



# Redução de Dimensionalidade

Cuidado para não ter uma visão parcial dos dados...



# Redução de Dimensionalidade

- Não é um problema trivial
- Em problemas reais, características discriminantes não são conhecidas a priori.
- Características raramente são totalmente independentes.
- Duas características irrelevantes, quando unidas pode formar uma nova característica relevante e com bom poder de discriminação
- Problema de otimização, em que o espaço de busca é  $2^N$  ( $N$  é o número de características).

# Redução de Dimensionalidade

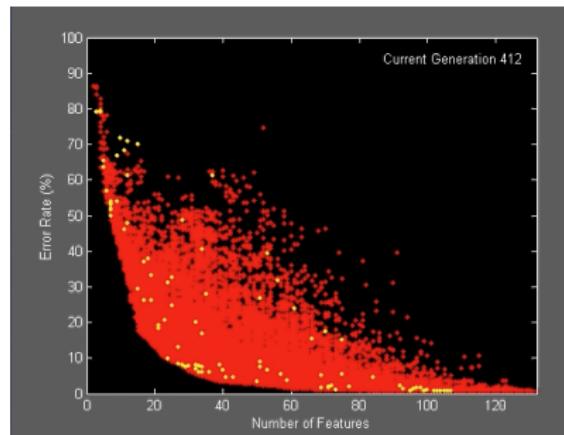
## Wrapper vs Filter

- Se o algoritmo de aprendizagem está envolvido no processo de seleção de atributos, o método é chamado de Wrapper
- Caso contrário, Filter.
- Os métodos wrapper geralmente fornecem melhores resultados, entretanto com um maior custo computacional.
  - ▶ O classificador é treinado para cada subconjunto de características que deve ser avaliado
  - ▶ Inviável para  $2^N$

# Meta Heurísticas

## Algoritmos Genéticos

- Devido ao poder de explorar grandes espaços de busca, algoritmos genéticos tem sido largamente utilizados em problemas de seleção de características
  - ▶ Um objetivo (desempenho ou um índice qualquer)
  - ▶ Múltiplos objetivos (quantidade de características, desempenho, etc..)
  - ▶ Pareto



## What else?

- Visualização (PCA, t-SNE)
- Ensembles
- Seleção Dinâmica de Classificadores
- Concept Drift
- Data Stream