APRENDIZADO DE MÁQUINA NA ERA DO BIG DATA

Eduardo Bezerra (CEFET/RJ)
ebezerra@cefet-rj.br

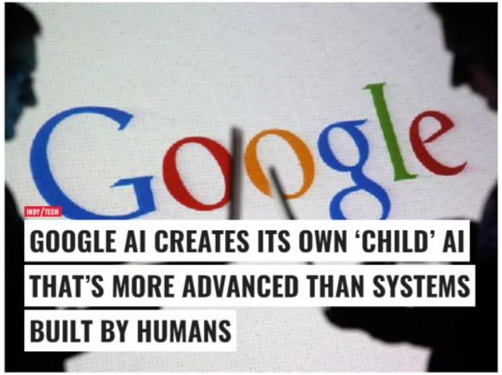
Visão Geral

- Introdução ao AM
- Deep Learning
- Considerações Finais

Aprendizado de Máquina

O que é Aprendizado de Máquina

[hype.]



Facebook's AI accidentally created its own language

by BRYAN CLARK — 9 months ago in ARTIFICIAL INTELLIGENCE

Machine Learning (to play Checkers)

Coined the term Machine Learning ("Field of study that gives computers the ability to learn without being explicitly programmed.")

"it will learn to play a better game of checkers than can be played by the person who wrote the program."

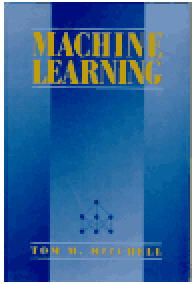
- search tree
- alpha-beta pruning
- scoring functions
- Minimax search
- TD-learning

1959

Arthur Samuel

O que é Aprendizado de Máquina

- Tom Mitchell (1998): "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

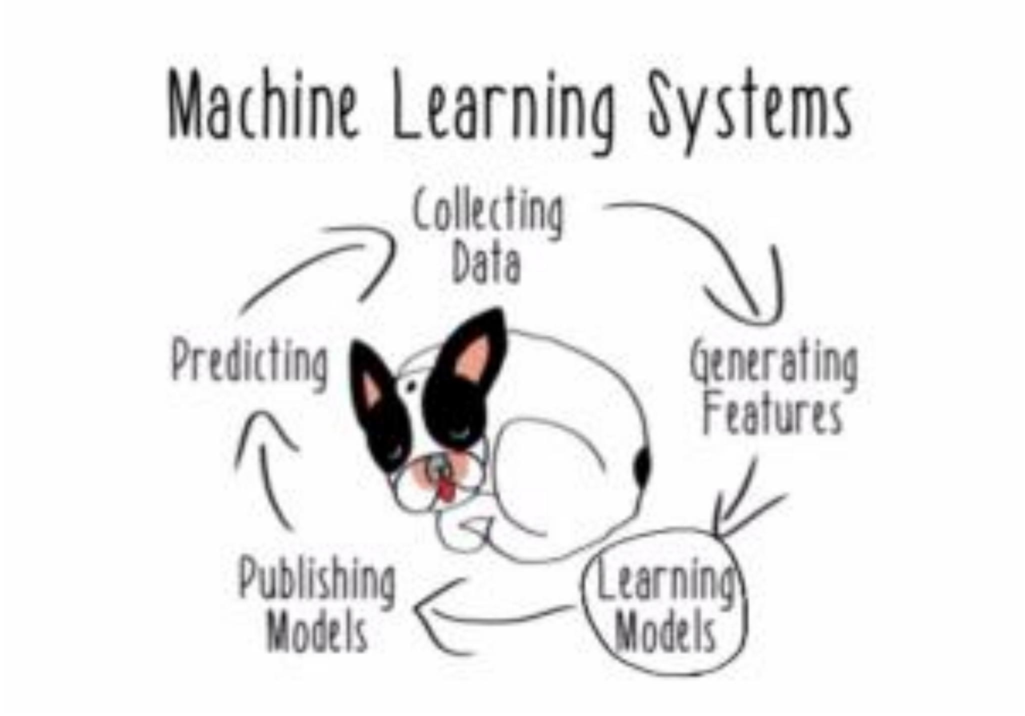


O que é Aprendizado de Máquina

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam.

- Components:
  - T
  - E
  - P

Machine Learning Systems



http://www.nosimpler.me/machine-learning/

Características (*features*)

Features:
1. Color: **Radish/Red**
2. Type : **Fruit**
3. Shape
etc...

Features:
1. Sky Blue
2. **Logo**
3. Shape
etc...

Features:
1. **Yellow**
2. **Fruit**
3. Shape
etc...

Terminologia

- *Training sample (training instance* or *training example).*
- Training set
- *Model*
- *Learning algorithm*

Tipos de Aprendizado de Máquina

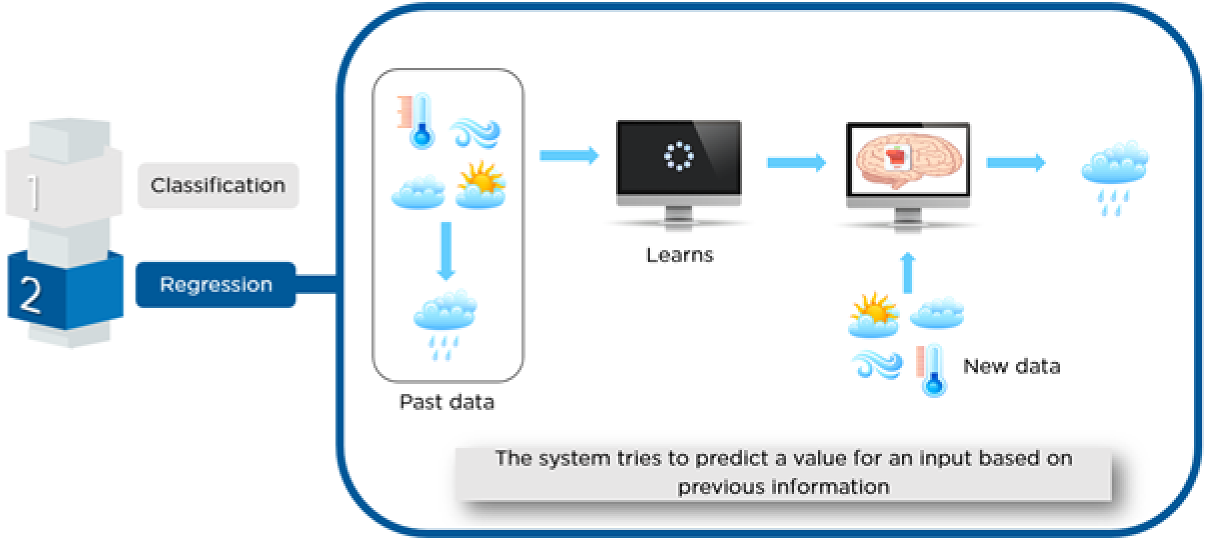
- Aprendizado supervisionado (*supervised learning*)
- Aprendizado não supervisionado (*unsupervised learning*)

Aprendizado Supervisionado

Aprendizado Supervisionado

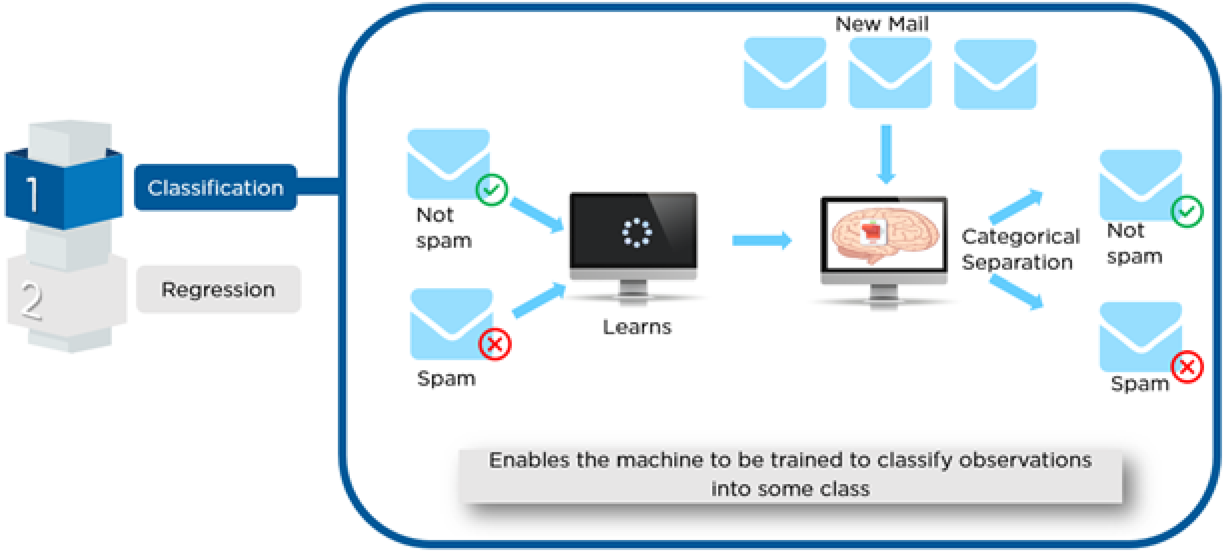
- O aprendiz (máquina) recebe as respostas corretas.
- Dois subtipos (tarefas):
  - Classificação: predizer valor discreto.
  - Regressão: predizer valor contínuo.

Regressão



https://www.quora.com/What-is-the-main-difference-between-classification-problems-and-regression-problems-in-machine-learning

Classificação



https://www.quora.com/What-is-the-main-difference-between-classification-problems-and-regression-problems-in-machine-learning

Detecção de spam

- Entrada: uma mensagem de e-mail
- Saída: spam/ham

- Construção:
  - Obter um coleção grande de mensagens como exemplos, cada uma rotulada como "spam" ou "ham"
  - Nota: alguém tem que rotular esses dados!
  - Objetivo: predizer o rótulo adequado para mensagens novas

- Características (*features*): os atributos usados para tomar a decisão (ham / spam)
  - As próprias palavras: FREE!
  - Padrões textuais: $dd, CAPS
  - Padrões não-textuais: SenderInContacts
  - …

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidencial and top secret. …

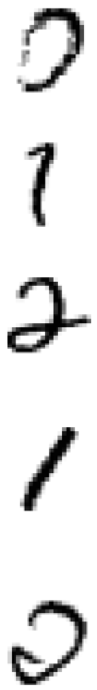
TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Reconhecimento de Dígitos

- Entrada: imagens / matrizes de pixels
- Saída: um dígito 0-9
- Construção:
  - Obter um coleção grande de mensagens (exemplos), cada uma rotulada com um dígito
  - Nota: alguém tem que rotular esses dados!
  - Objetivo: predizer o rótulo adequado para imagens novas
- Características: os atributos usados para tomar a decisão
  - Pixels: (6,8)=ON
  - Padrões de forma: NumComponents, AspectRatio, NumLoops
  - …



0

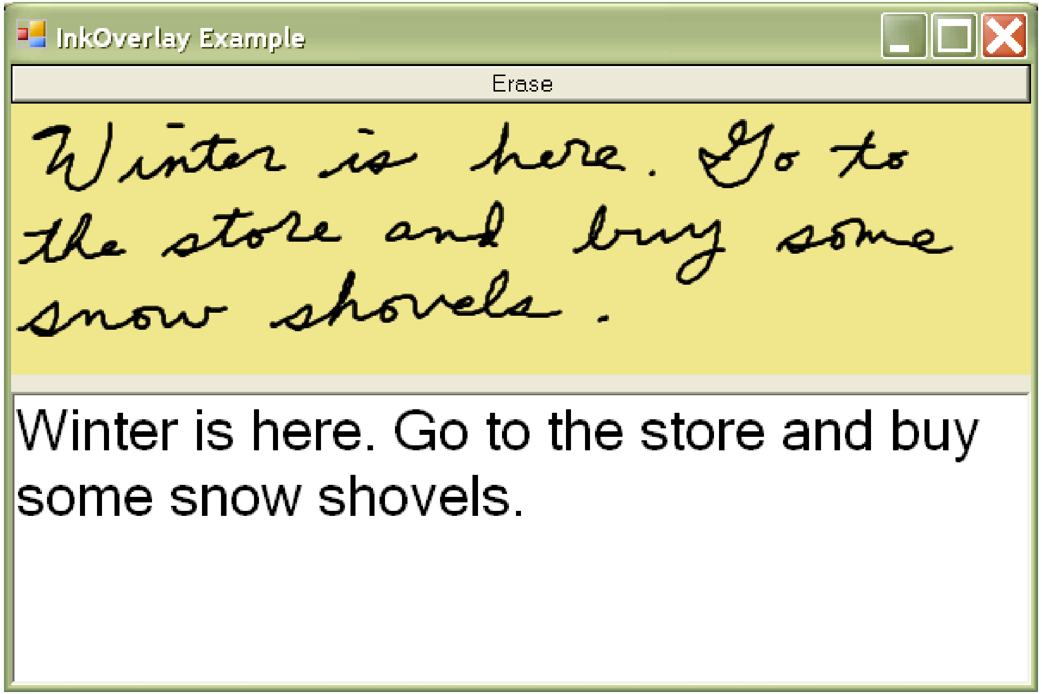
1

2

1

??

Reconhecimento de Caracteres



Tradução Automática

https://intlcontact.com/category/services/translation-service/

Análise de Sentimentos

Nome

Endereço

Cidade

UF

CEP

Modelo

BMX 330
BMX 550
RBX 12
Sirax 220
Street E3

Descrição do problema

Processamento de Linguagem Natural

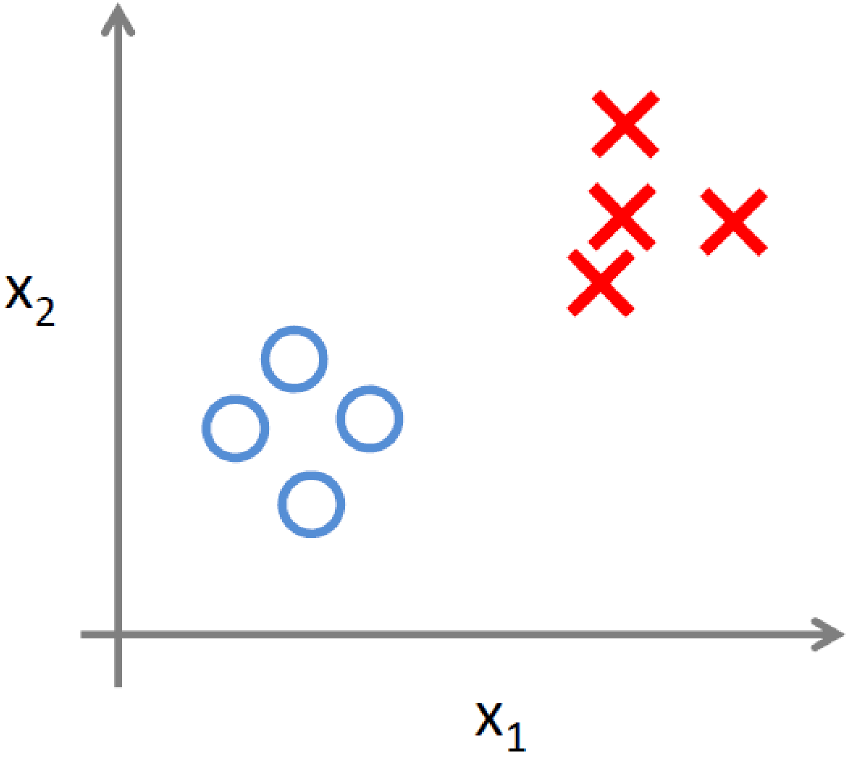
Classificação: Outras Aplicações

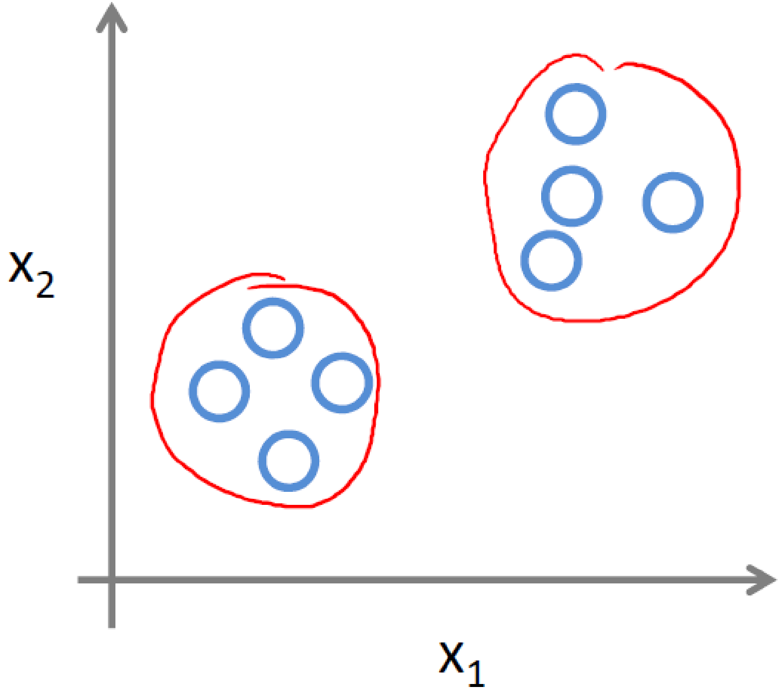
- Classificação: dados objetos de entrada, predizer rótulos.
- Exemplos:
  - OCR (entrada: imagens, classes: caracteres)
  - Diagnose médica (entrada : sintomas, classes: doenças)
  - Correção automática de redações (entrada : documentos, classes: notas)
  - Detecção de fraude (entrada : atividades na conta, classes: fraude / legítimo)
  - Roteamento de notícias
  - … muitos mais

- Classificação é uma tecnologia importante comercialmente!

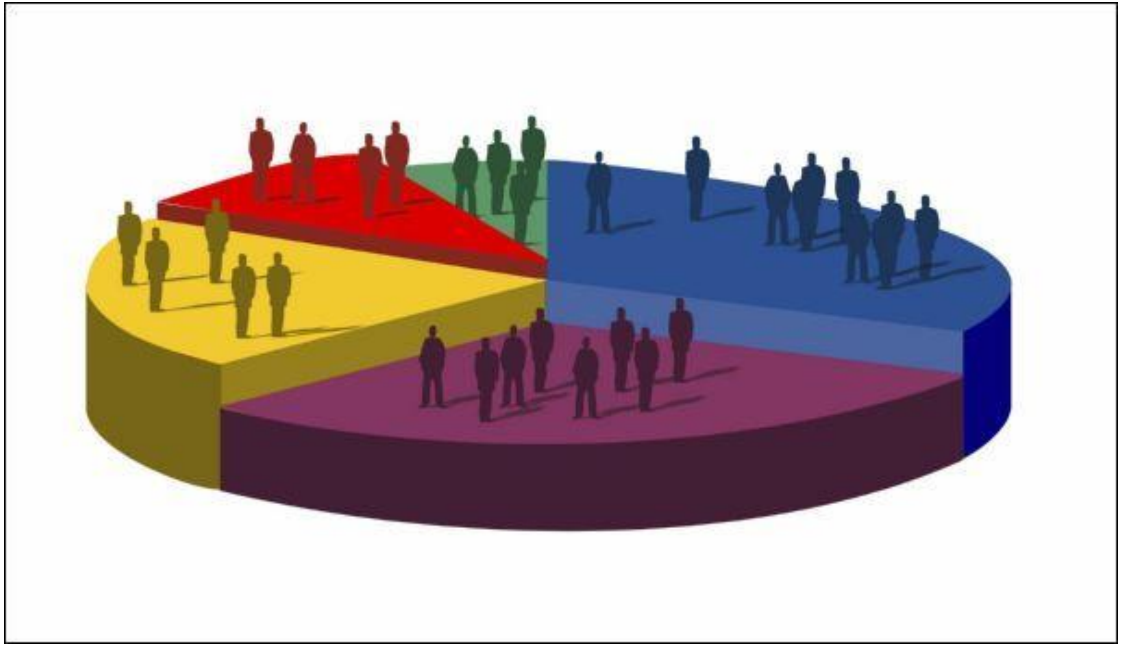
Aprendizado Não Supervisionado

Aprendizado Não Supervisionado

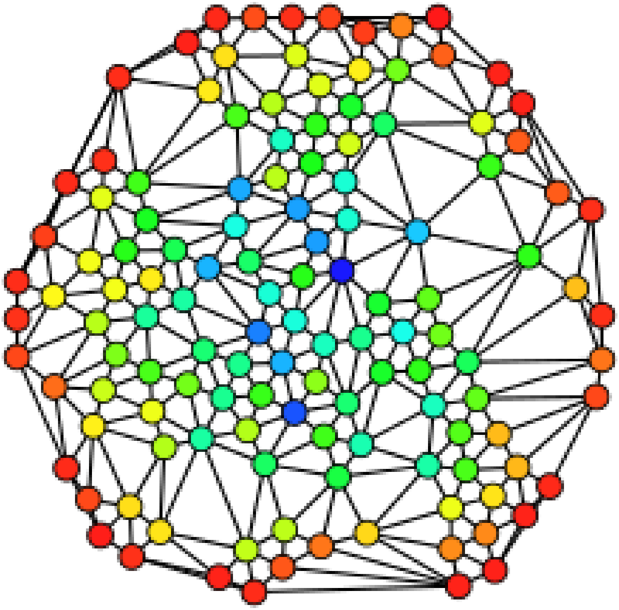


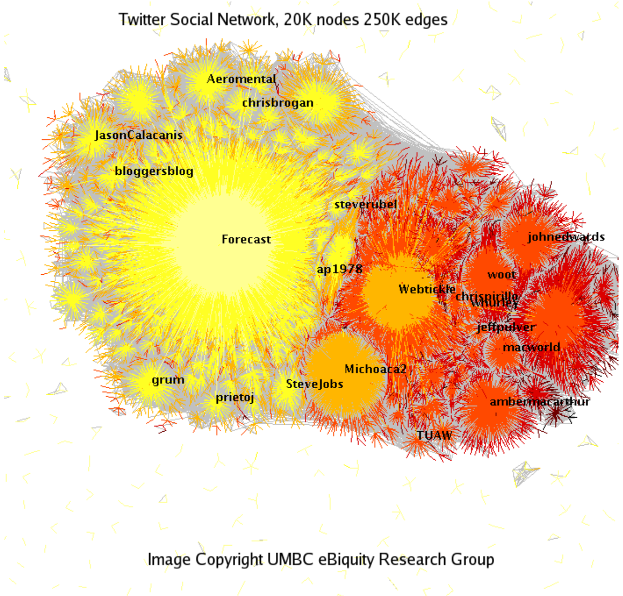


Segmentação de mercado (CRM)



Análise de Redes Sociais





Detecção de Valores Extremos *(outlier detection)*

Deep Learning

Neural Nets Renaissance

2006

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

Learning multiple layers of representation

2007

Geoffrey E. Hinton

Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, M5S 3G4, Canada

2009

ImageNet: A Large-Scale Hierarchical Image Database

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei
Dept. of Computer Science, Princeton University, USA
{jiadeng, wdong, rsocher, jial, li, feifeili}@cs.princeton.edu

2009

Learning Deep Architectures for AI

Yoshua Bengio

Large-scale Deep Unsupervised Learning using Graphics Processors

2009

Rajat Raina RAJATR@CS.STANFORD.EDU
Anand Madhavan MANAND@STANFORD.EDU
Andrew Y. Ng ANG@CS.STANFORD.EDU
Computer Science Department, Stanford University, Stanford CA 94305 USA

2000s

Deep Learning Explosion

2010s

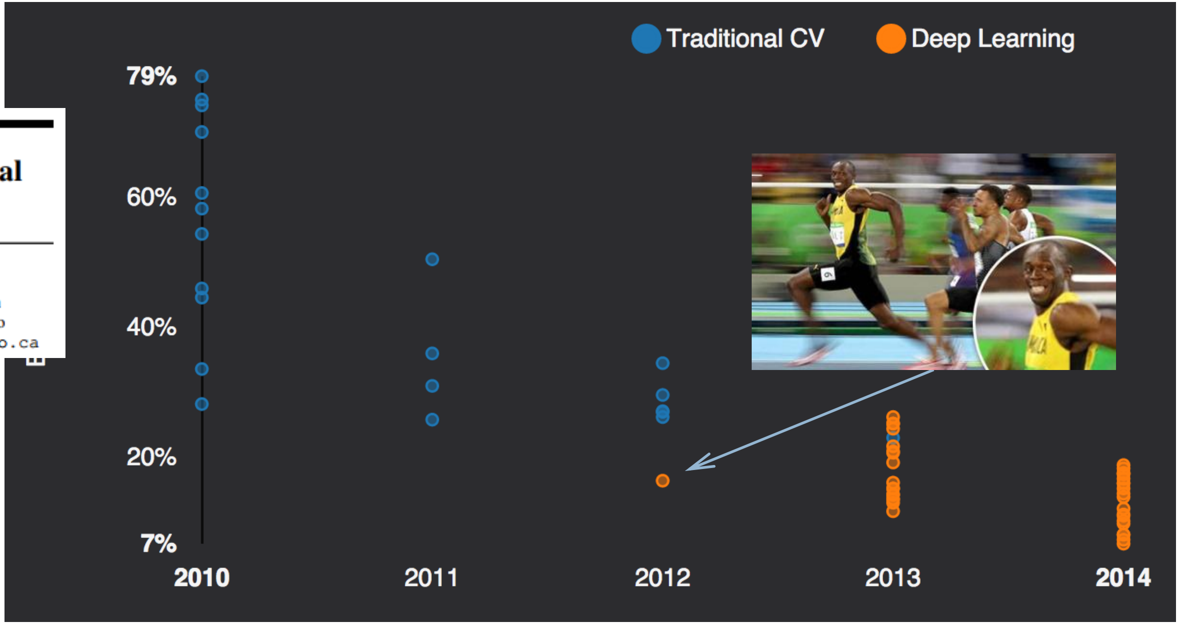
Object Detection (in Images)

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



Credits: Mathew Zeiler (Clarifai)

2012

Speech Recognition

Method	PER
CD-HMM [26]	27.3%
Augmented conditional Random Fields [26]	26.6%
Randomly initialized recurrent Neural Nets [27]	26.1%
Bayesian Triphone GMM-HMM [28]	25.6%
Monophone HTMs [29]	24.8%
Heterogeneous Classifiers [30]	24.4%
Monophone randomly initialized DNNs (6 layers)[13]	23.4%
Monophone DBN-DNNs (6 layers) [13]	22.4%
Monophone DBN-DNNs with MMI training [31]	22.1%
Triphone GMM-HMMs discriminatively trained w/ BMMI [32]	21.7%
Monophone DBN-DNNs on fbank (8 layers) [13]	20.7%
Monophone mcRBM-DBN-DNNs on fbank (5 layers) [33]	20.5%
Monophone convolutional DNNs on fbank (3 layers) [34]	20.0%

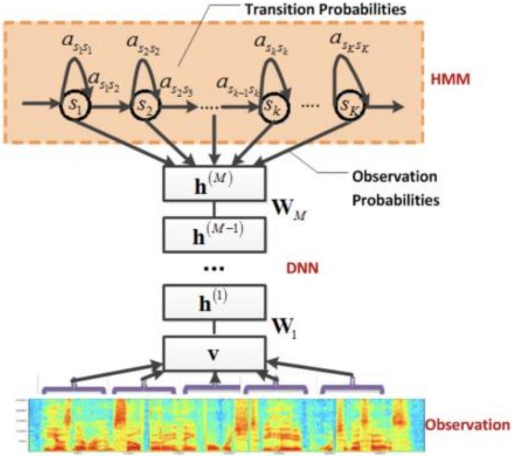


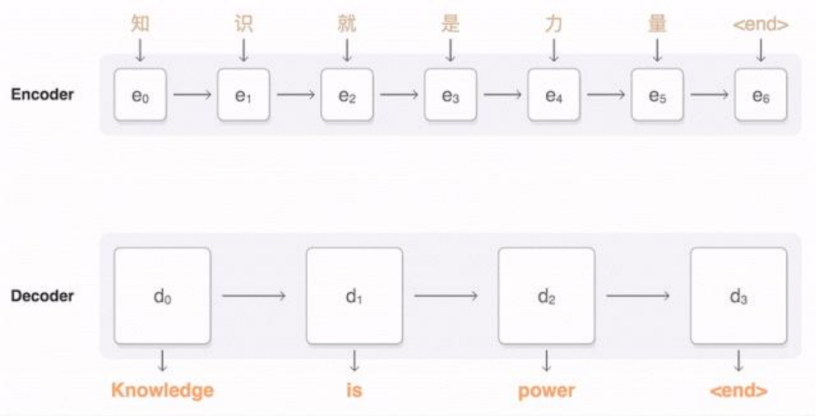
Fig. 1. Diagram of our hybrid architecture employing a deep neural network. The HMM models the sequential property of the speech signal, and the DNN models the scaled observation likelihood of all the senones (tied tri-phone states). The same DNN is replicated over different points in time.

Deep Neural Networks for Acoustic Modeling in Speech Recognition

Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury

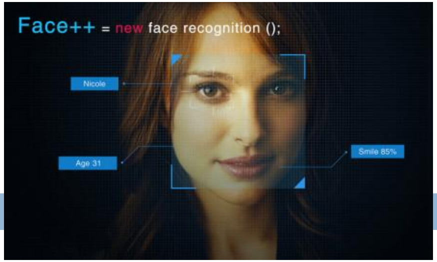
2012

Language Translation

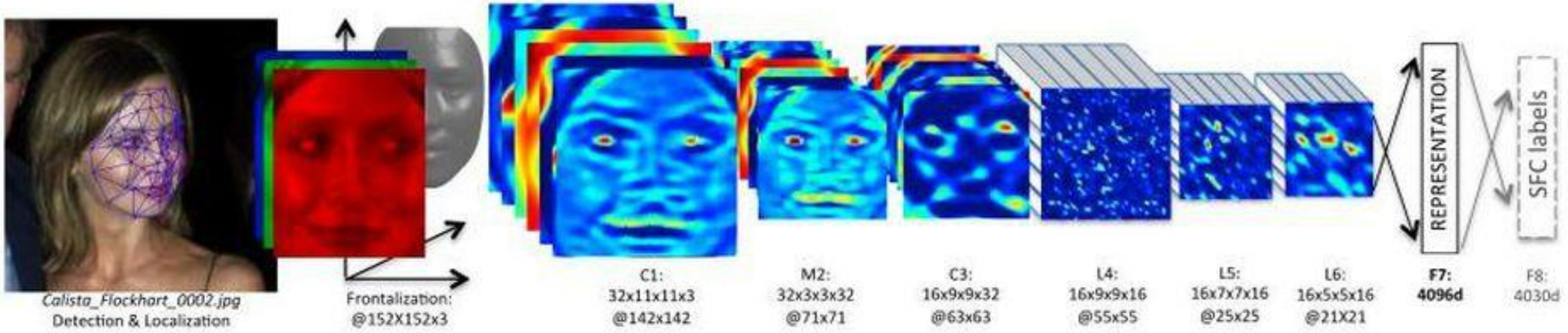


2014

Face Recognition



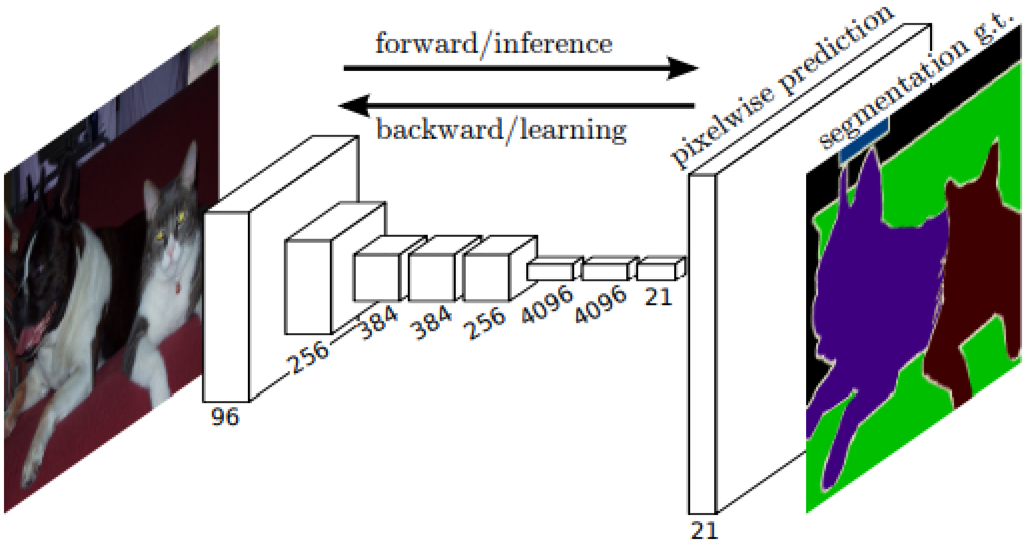
DeepFace: Our method reaches an accuracy of 97.35% […], reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.



2014

https://research.fb.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/

Semantic Segmentation

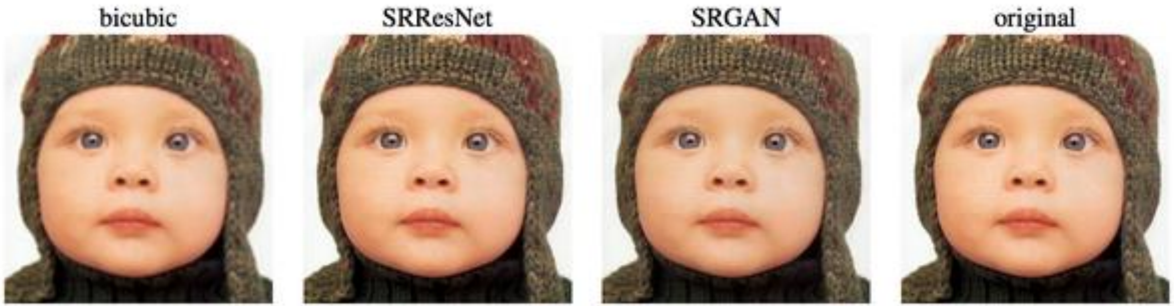


Fully Convolutional Networks for Semantic Segmentation

Jonathan Long* Evan Shelhamer* Trevor Darrell
UC Berkeley
{jonlong,shelhamer,trevor}@cs.berkeley.edu

2014

Image Generation/Superresolution

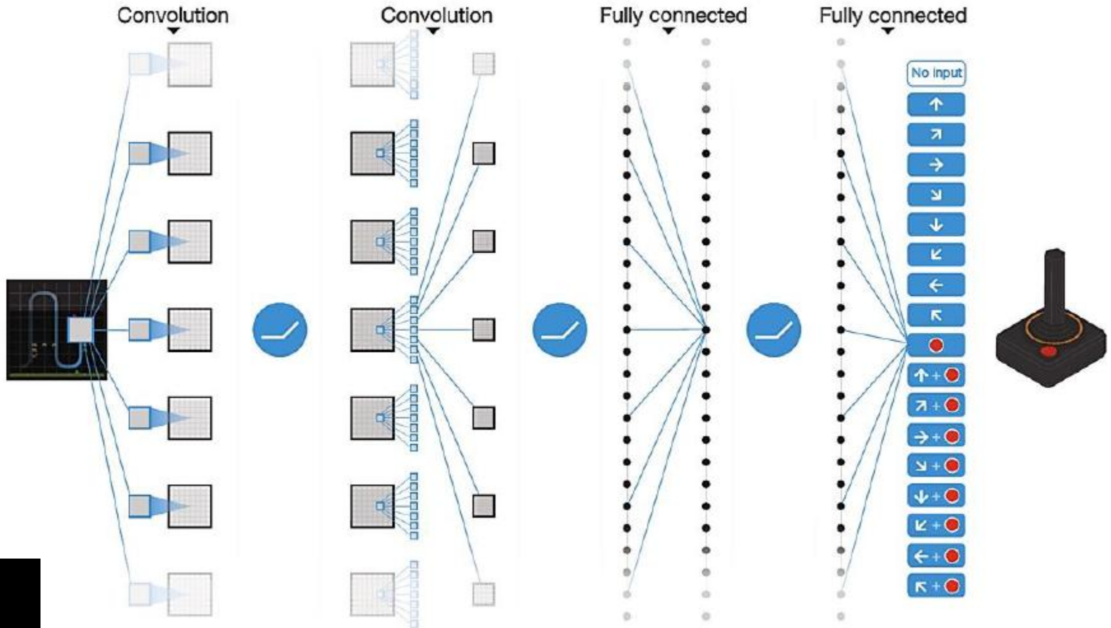


2014

Generative Adversarial Nets

Deep Reinforcement Learning

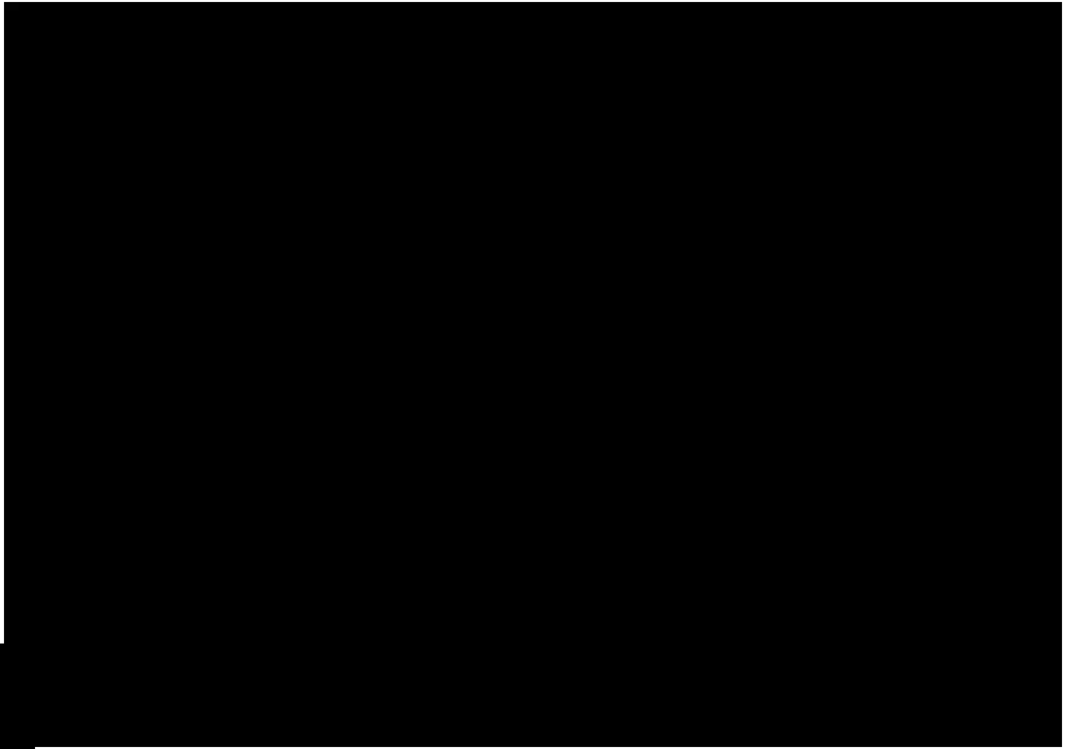
Deep Q-Learning



2015

Games

Deep Reinforcement Learning



2015

Games

Deep Reinforcement Learning

2016

AlphaGo

Automatic Labelling

A close up of a hot dog on a bun.

A bath room with a toilet and a bath tub.

A vase filled with flower sitting on a table.

2016

Long-term Recurrent Convolutional Networks for Visual Recognition and Description, 2016.


DEMO

Considerações Finais

Machine Learning: success factors

- Big Data
  - 1980s: MNIST ~ 70k
  - 2010s: ImageNet ~ 10^6
- Hadware improvements
- Crowdsourcing

"What was wrong in the 80's is that we didn't have enough data and we didn't have enough computer power"

Geoffrey Hinton

ML & Big Data

- Atualmente, é relativamente fácil obter conjuntos de dados de treinamento da ordem de 10^6 exemplos:
  - e-commerce portals
  - Kaggle (https://www.kaggle.com)
  - IoT
  - ...

ML & Big Data

- Quanto mais dados, melhor?

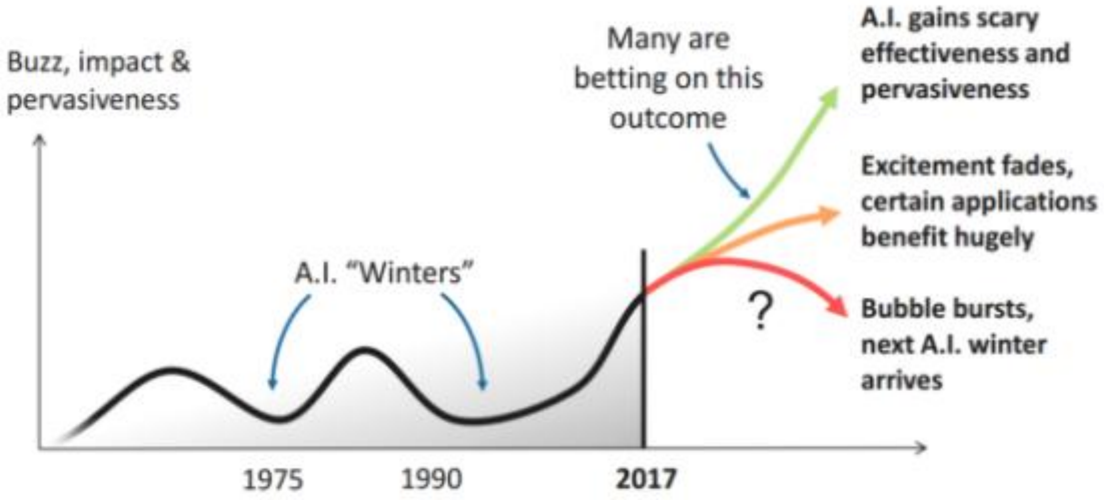
  "[…] what we're seeing consistently is that the bigger you can run these models, the better they perform. If you train one of these algorithms on one computer, you know, it will do pretty well. If you train them on 10, it will do even better. If you train on 100, even better. And we found that when we trained it on 16,000 CPU cores, […], that was the best model we were able to train."

  --Andrew Ng

http://www.npr.org/2012/06/26/155792609/a-massive-google-network-learns-to-identify

But …

Is Winter Coming Again?!

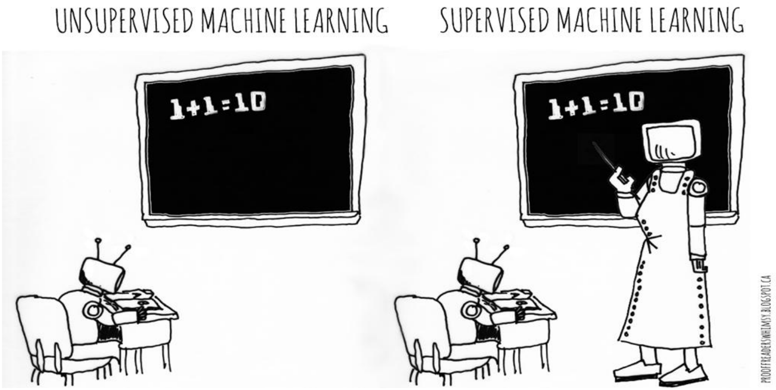


Credits: Monty Barlow

Unsupervised Learning

Current models are hungry for labeled data.

Today's DL is supervised learning.



"The Revolution Will Not be Supervised." –Yann Lecun

Deep RL Takes Too Long to Train

RL systems require a gazillion trials!

Natural Language Understanding

- Headlines:
  - Enraged Cow Injures Farmer With Ax
  - Hospitals Are Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor's Desk
  - Iraqi Head Seeks Arms
  - Local HS Dropouts Cut in Half
  - Juvenile Court to Try Shooting Defendant
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks

- Why are these funny?

Source: CS188

Common Sense Knowledge

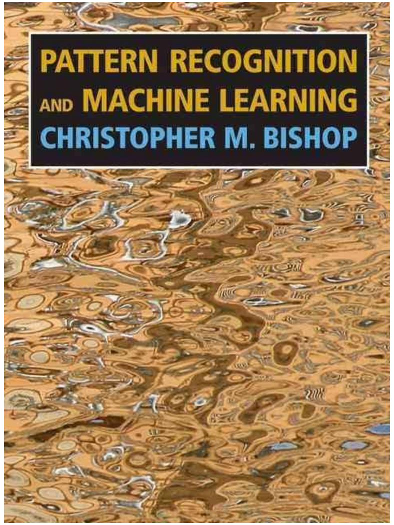
"If a mother has a son, then the son is younger than the mother and remains younger for his entire life."

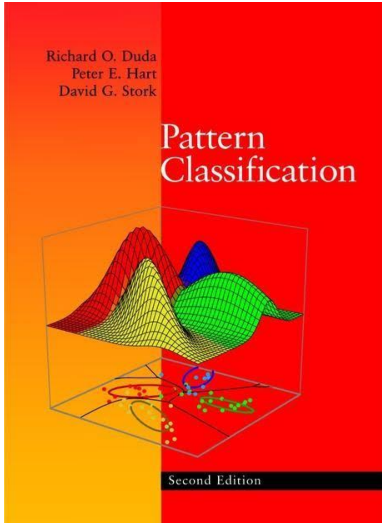
"If President Trump is in Washington, then his left foot is also in Washington,"

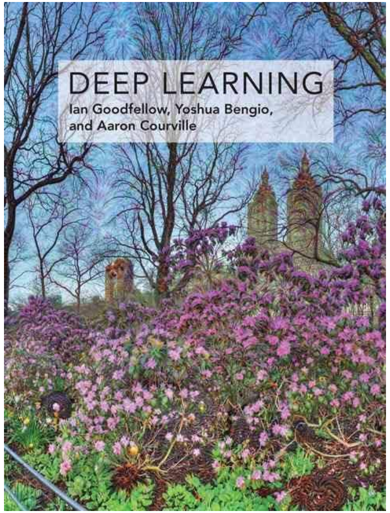
Para Saber Mais – Curso Online

- Machine Learning
  - https://www.coursera.org/learn/machine-learning
- Deep Learning Specialization
  - https://www.coursera.org/specializations/deep-learning
- Convolutional Neural Networks for Visual Recognition
  - http://cs231n.stanford.edu/

Para Saber Mais – Livros







PPCIC – CEFET/RJ
Programa de Pós-Graduação em Ciência da Computação
http://eic.cefet-rj.br/ppcic

APRENDIZADO DE MÁQUINA NA ERA DO BIG DATA

OBRIGADO!

Eduardo Bezerra (ebezerra@cefet-rj.br)

BACKUP SLIDES

MapReduce

MapReduce

- Alguns problemas de AM podem ser muito grandes para resolver em apenas uma máquina.
- Uma abordagem geral para distribuir processamento é denominada MapReduce.
  - Aplicável a alguns algoritmos de AM...

Jeff Dean Sanjay Ghemawat

MapReduce

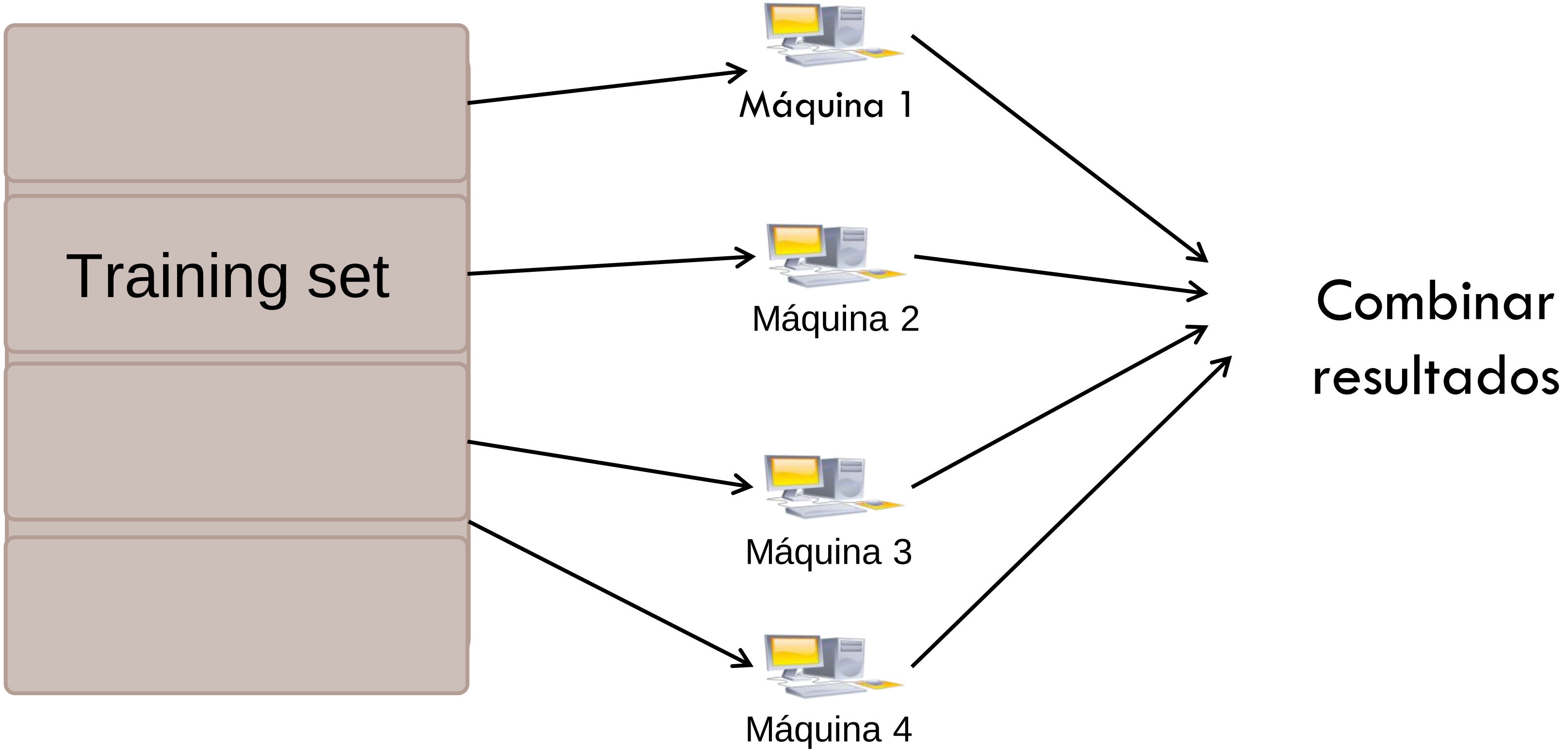
Muitos algoritmos de AM podem ser expressados como a computação de somatórios de funções sobre o conjunto de treinamento.

e.g., para a regressão logística, precisamos de:

$$J_{train}(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) - (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))$$

$$\frac{\partial}{\partial \theta_j} J_{train}(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

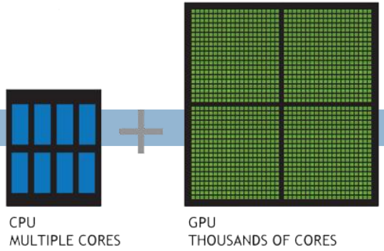
MapReduce

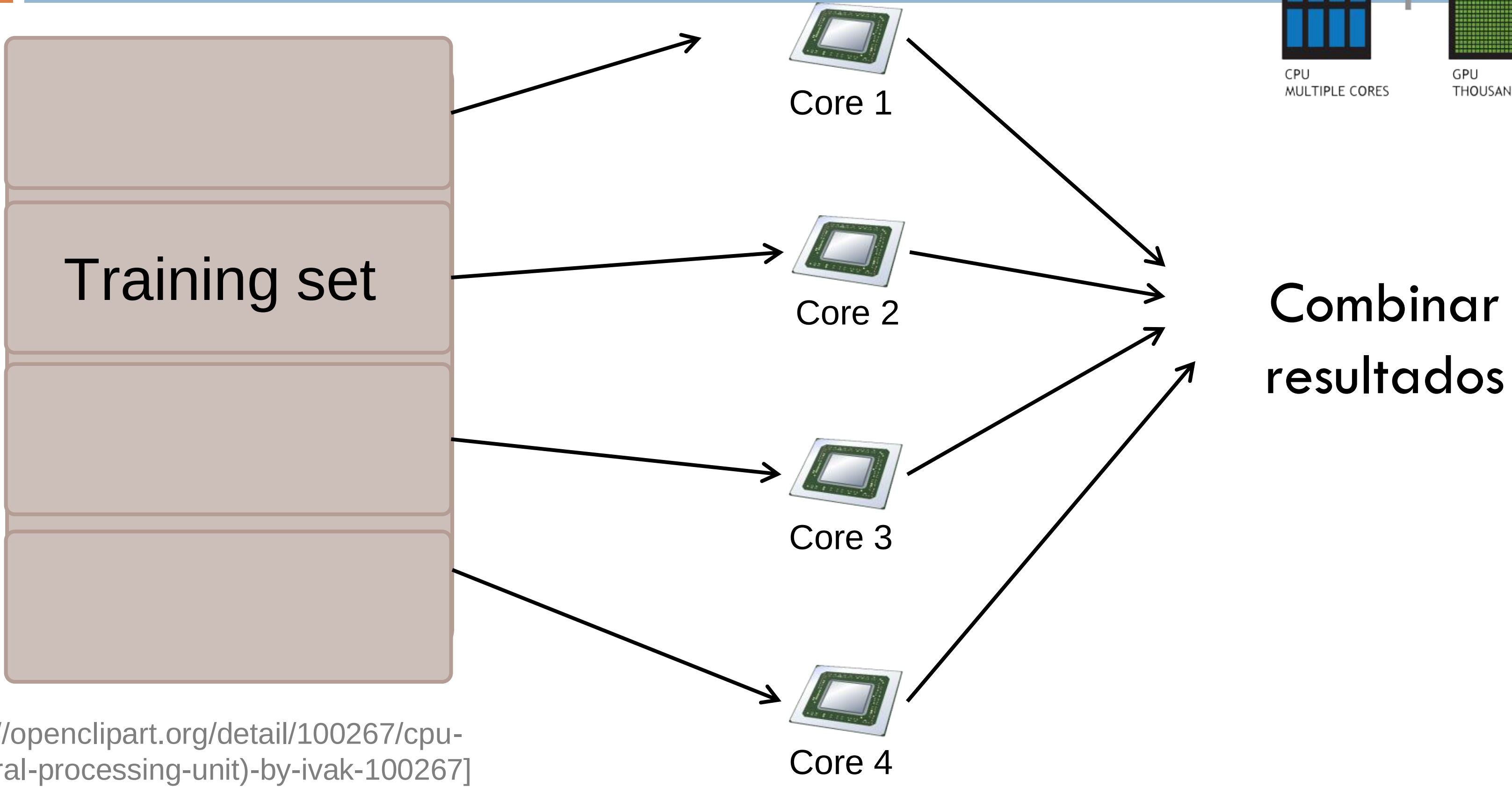


[http://openclipart.org/detail/17924/computer-by-aj]

MapReduce

Máquinas multi-core & GPUs

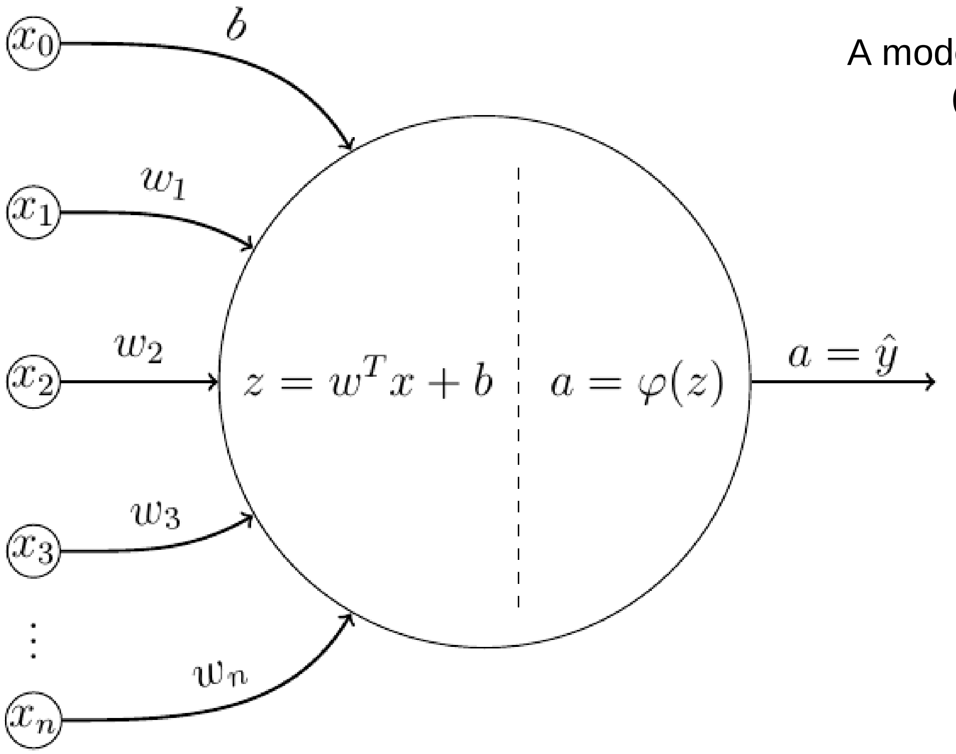




[http://openclipart.org/detail/100267/cpu-(central-processing-unit)-by-ivak-100267]

Artificial Neural Nets (ANNs)

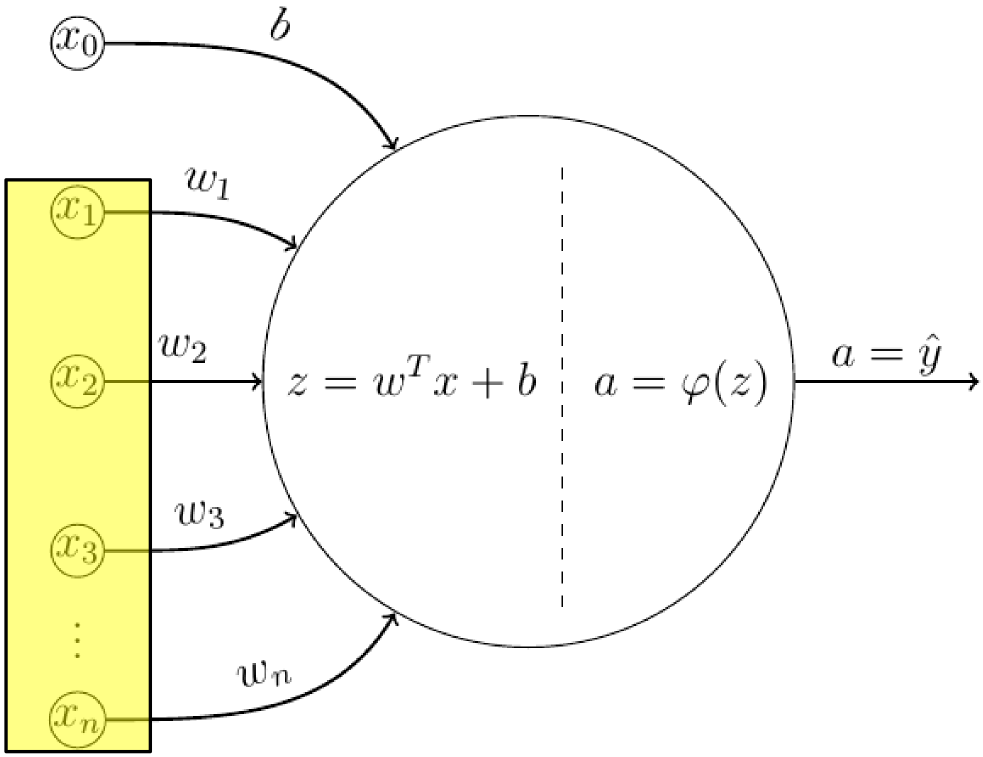
Artificial Neuron



A model inspired in the real one (biological neuron).

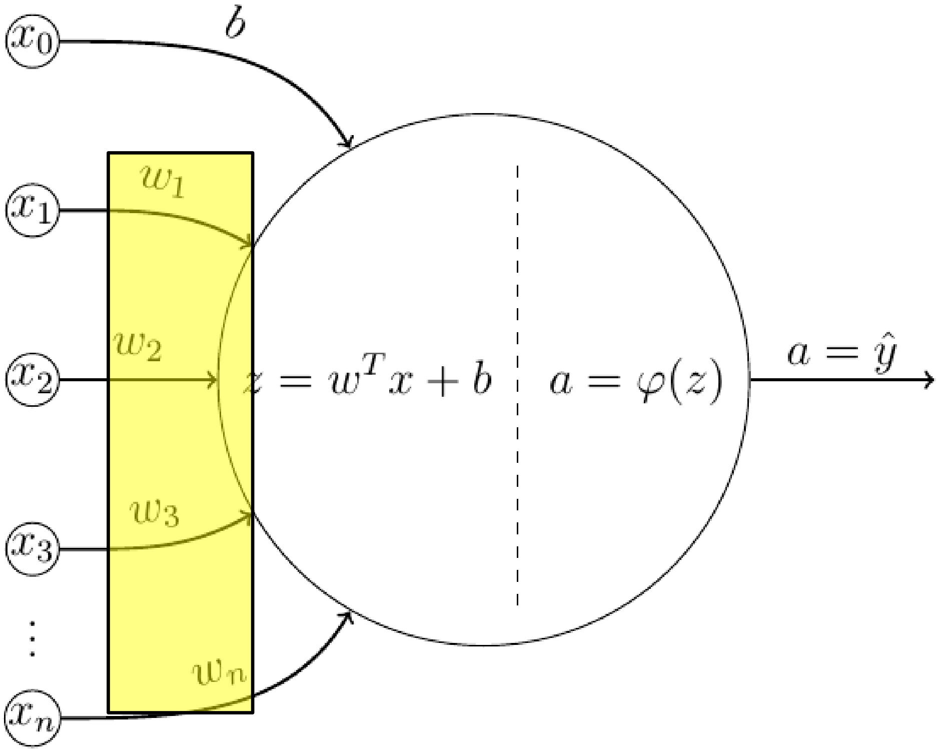
Artificial Neuron - input

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

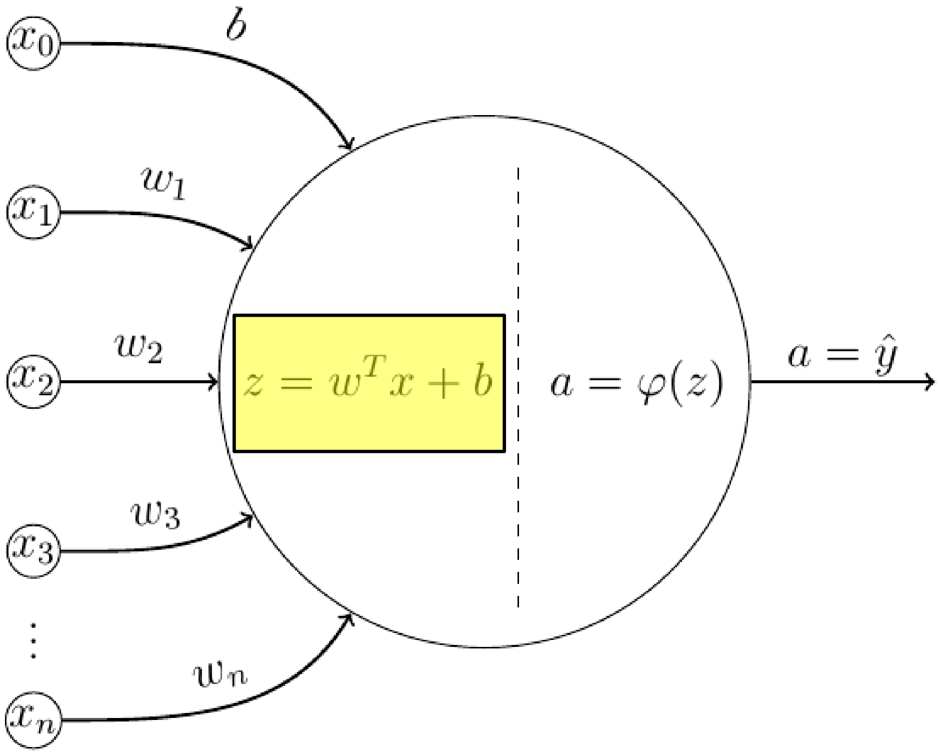


Artificial Neuron – parameters

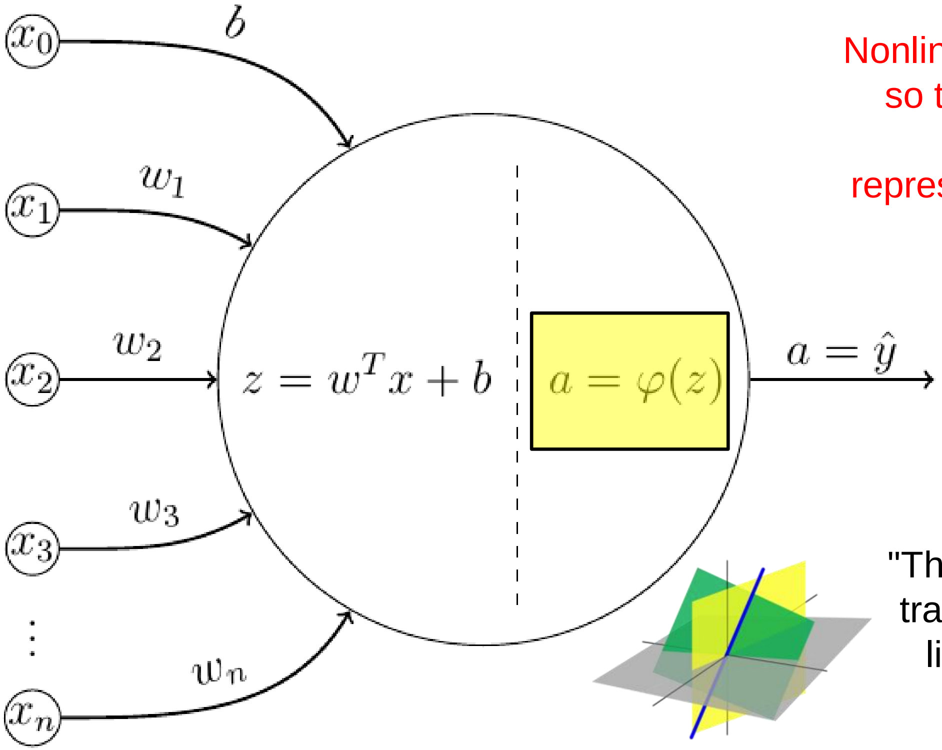
$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$



Artificial Neuron – pre-activation



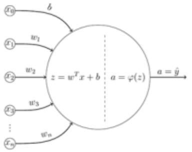
Artificial Neuron – activation function

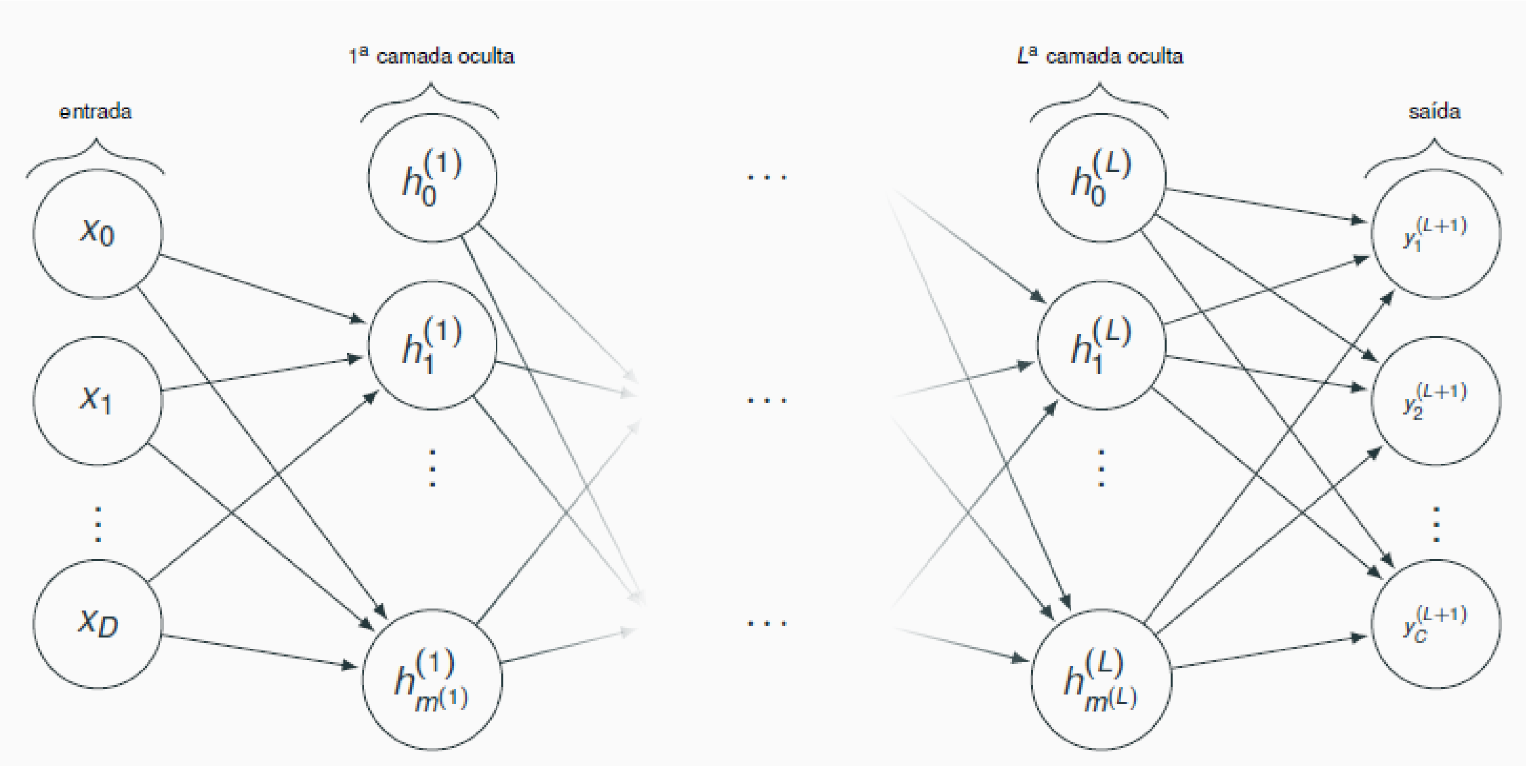


Nonlinearities are necessary so that the network can learn complex representations of the data.

"The composition of linear transformations is also a linear transformation"

Artificial Neural Net





Feedforward Neural Network