Deep Learning Meets Big Data

Nivio Ziviani

Kunumi

Professor Emérito do DCC/UFMG

Membro da Academia Brasileira de Ciências


Deep Learning
É uma área da Computação Cognitiva que usa e desenvolve algoritmos que permitem modelar abstrações de alto nível em uma grande base de dados. Esses dados são processados em múltiplas camadas -por isso- "Deep Learning".

Reconhecimento de objetos em tempo real

Source: “Delving Deep into Rectifiers:
Surpassing Human-Level Performance on ImageNet Classification"
Microsoft Research http://arxiv.org/pdf/1502.01852v1.pdf

Descrição de Imagem

two pizzas on the stove, one with mushrooms and the other with basil

Fonte: deeplearning.cs.toronto.edu/i2t

Descrição de Imagem

huge crowd are gathered probably to demonstrate

Fonte: deeplearning.cs.toronto.edu/i2t

Aprendizado Dinâmico

Source: Playing Atari with deep reinforcement learning
https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf

Before training

peaceful swimming

Deep Learning

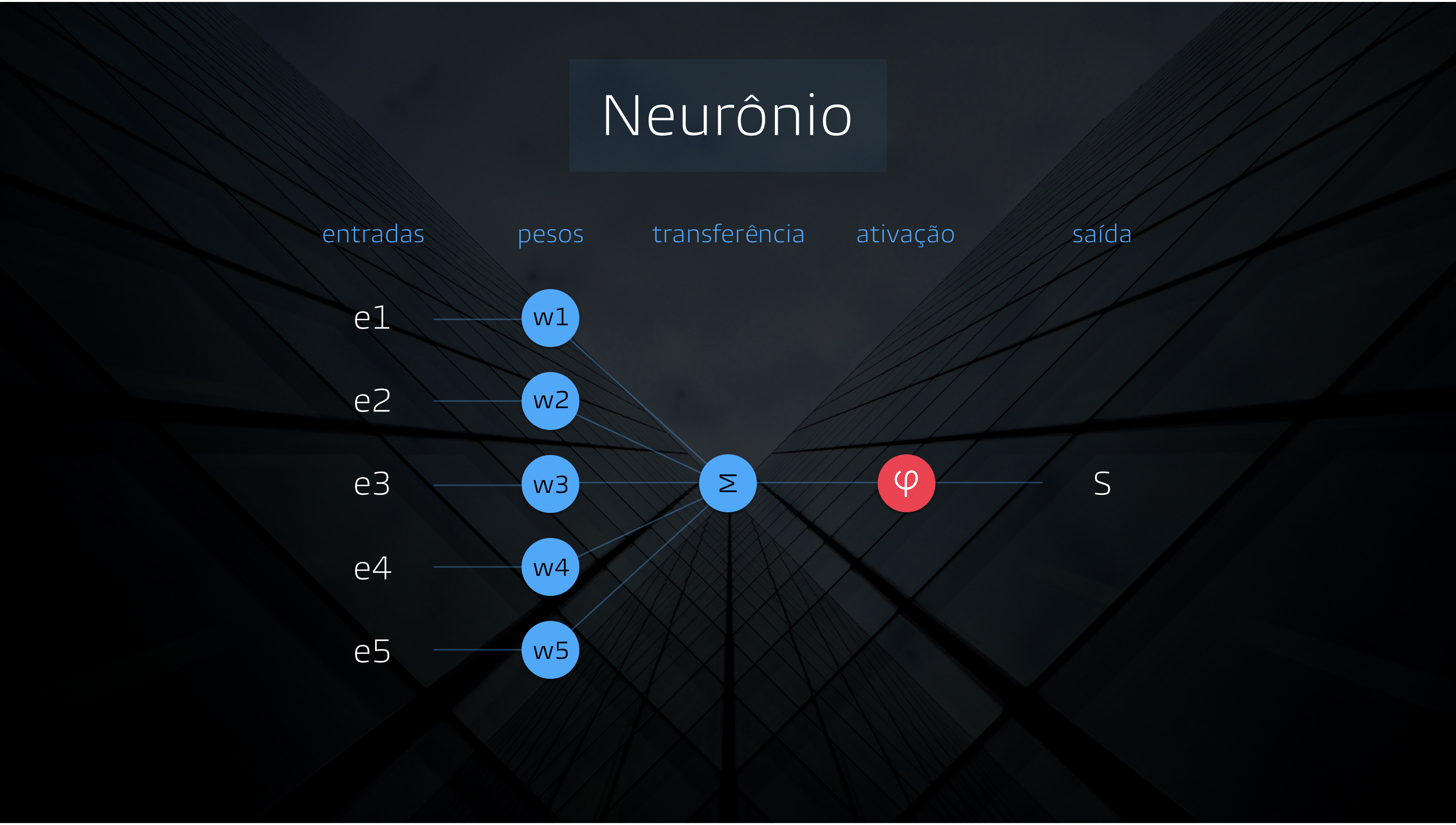
Redes profundas são baseadas no aprendizado da representação de dados. Por exemplo, palavras são representadas por vetores de números reais.

Deep Learning

Aprendizado pode ser supervisionado (classificação) ou não-supervisionado (análise de padrões).

Trabalha bem com dados estruturados e não estruturados


Deep Learning
Camadas de características usam um procedimento
*que **não é projetado por pessoas**: elas são aprendidas usando*
um procedimento de aprendizado de propósito geral !


Neurônio
entradas
pesos
transferência
ativação
saída
e1
e2
e3
e4
e5
w1
w2
w3
w4
w5
Σ
φ
S

O que aprende uma rede neuronal?

imagem
original

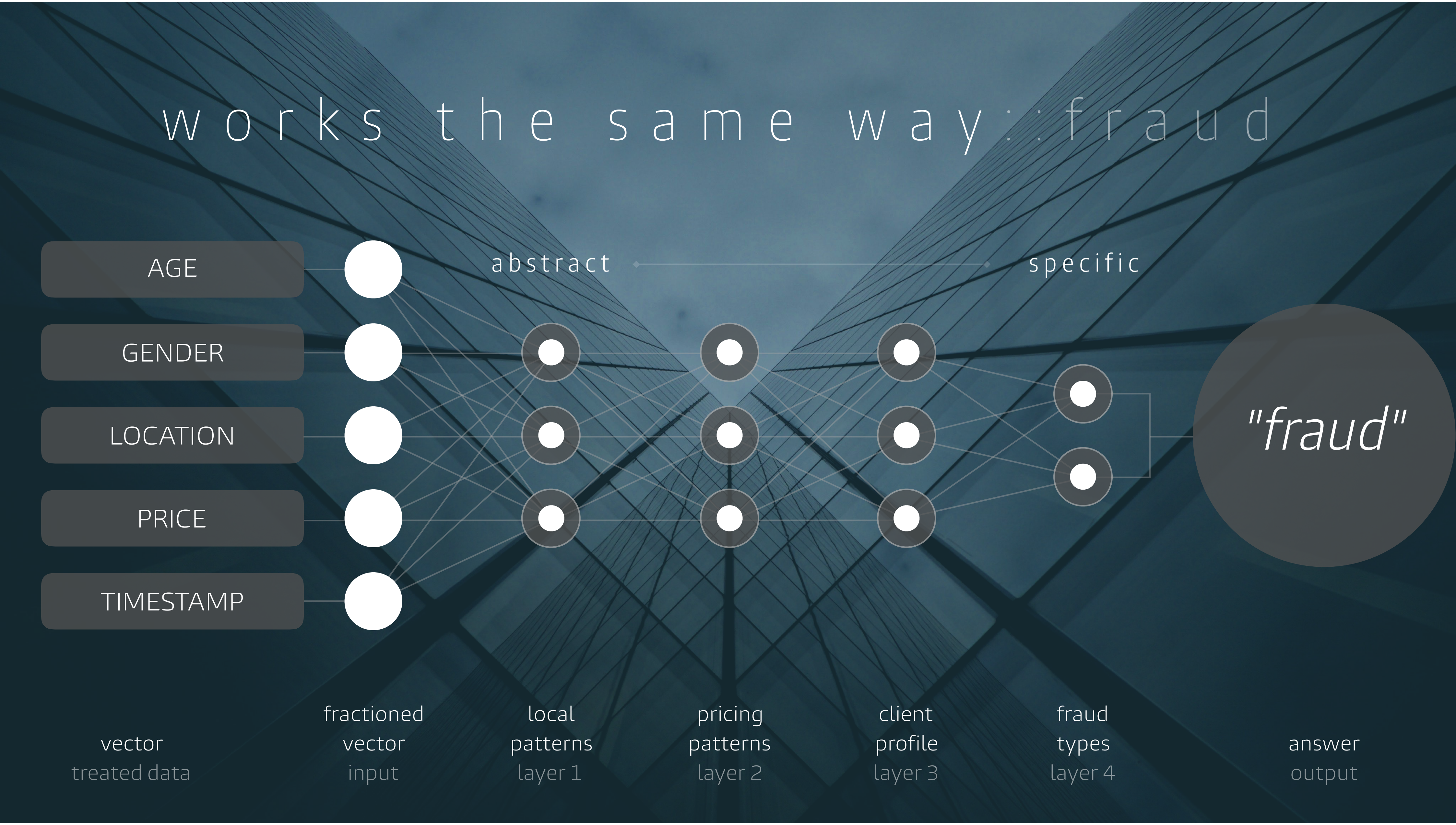
cada ponto
é avaliado

0.752	0.601	0.822
0.537	0.013	0.673
0.799	0.498	0.982

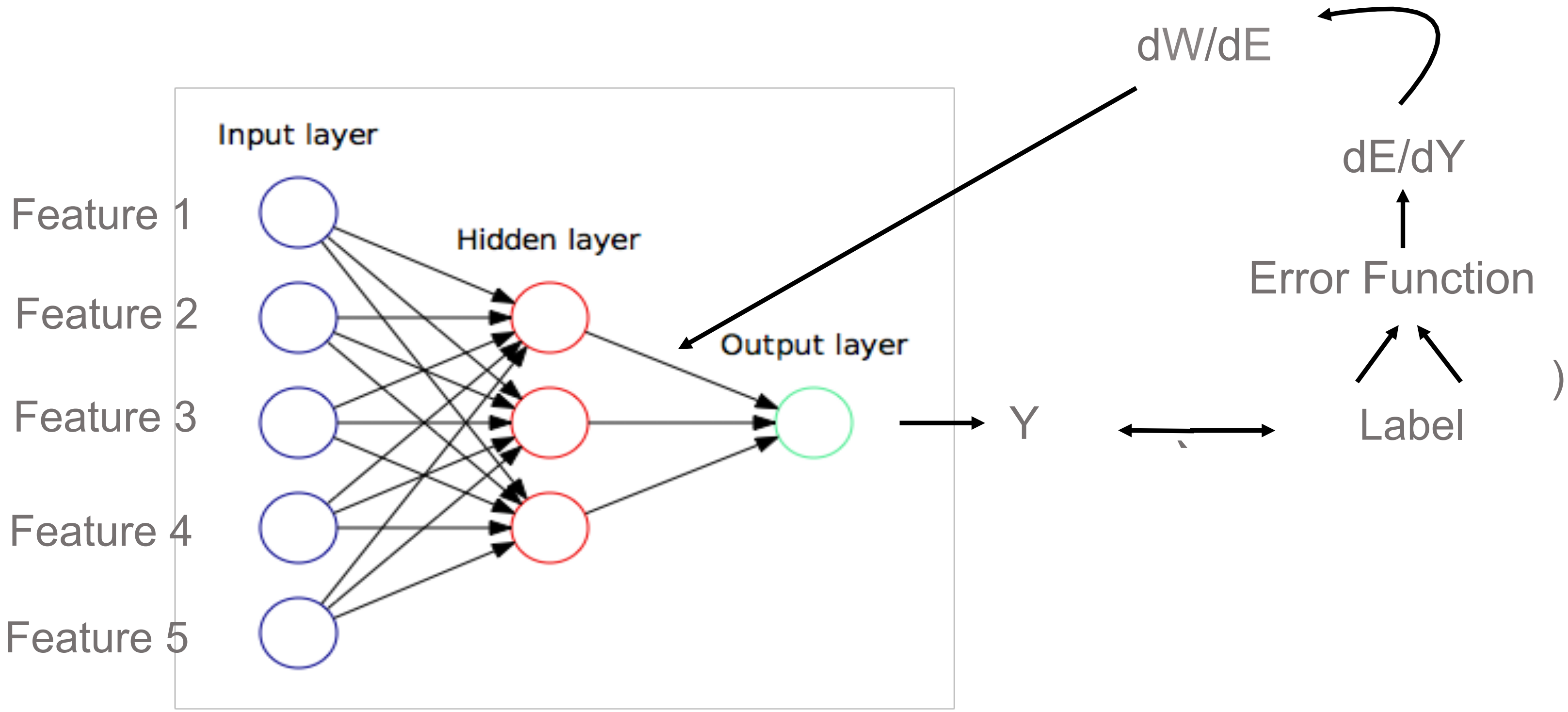
transformado num
número entre Ø e 1

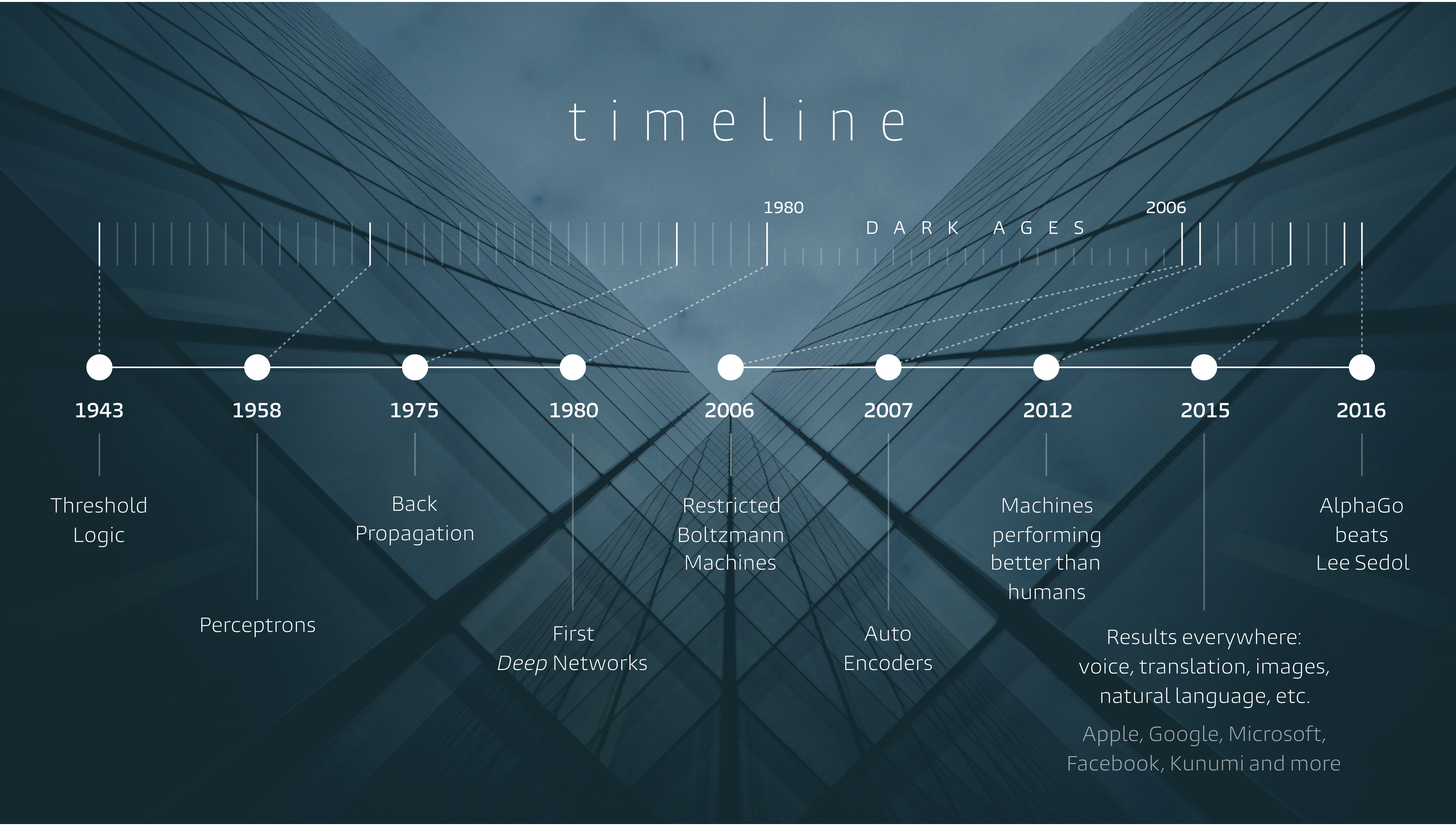
e reagrupados
formam o vetor

O que aprende uma rede neuronal?


works the same way : fraud
abstract
specific
AGE
GENDER
LOCATION
PRICE
TIMESTAMP
"fraud"
vector
treated data
fractioned
vector
input
local
patterns
layer 1
pricing
patterns
layer 2
client
profile
layer 3
fraud
types
layer 4
answer
output

Neural Networks: Training

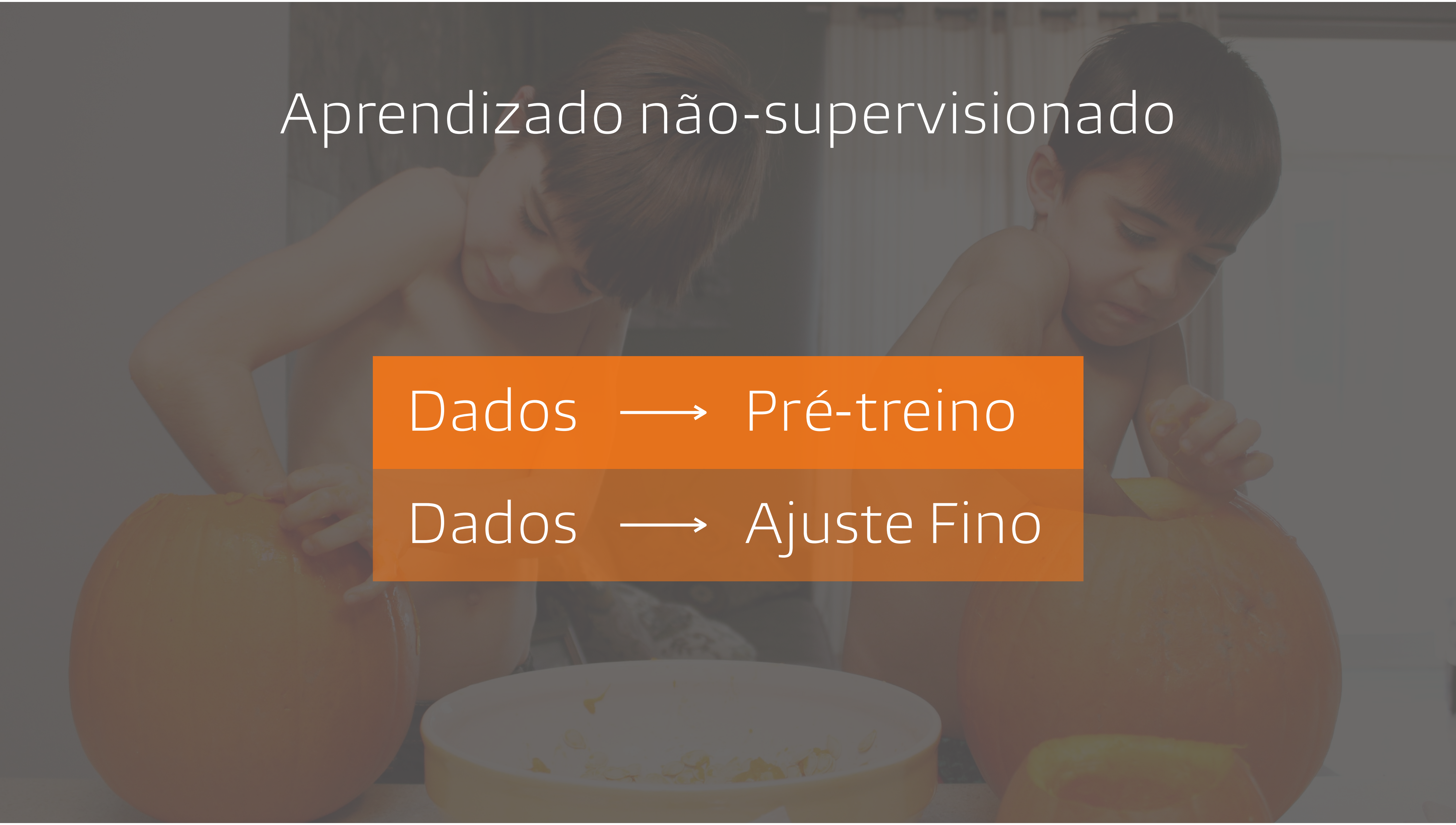



timeline
1980
DARK AGES
2006
1943
Threshold Logic
1958
Perceptrons
1975
Back Propagation
1980
First *Deep* Networks
2006
Restricted Boltzmann Machines
2007
Auto Encoders
2012
Machines performing better than humans
2015
Results everywhere: voice, translation, images, natural language, etc.
Apple, Google, Microsoft, Facebook, Kunumi and more
2016
AlphaGo beats Lee Sedol

Aprendizado não-supervisionado

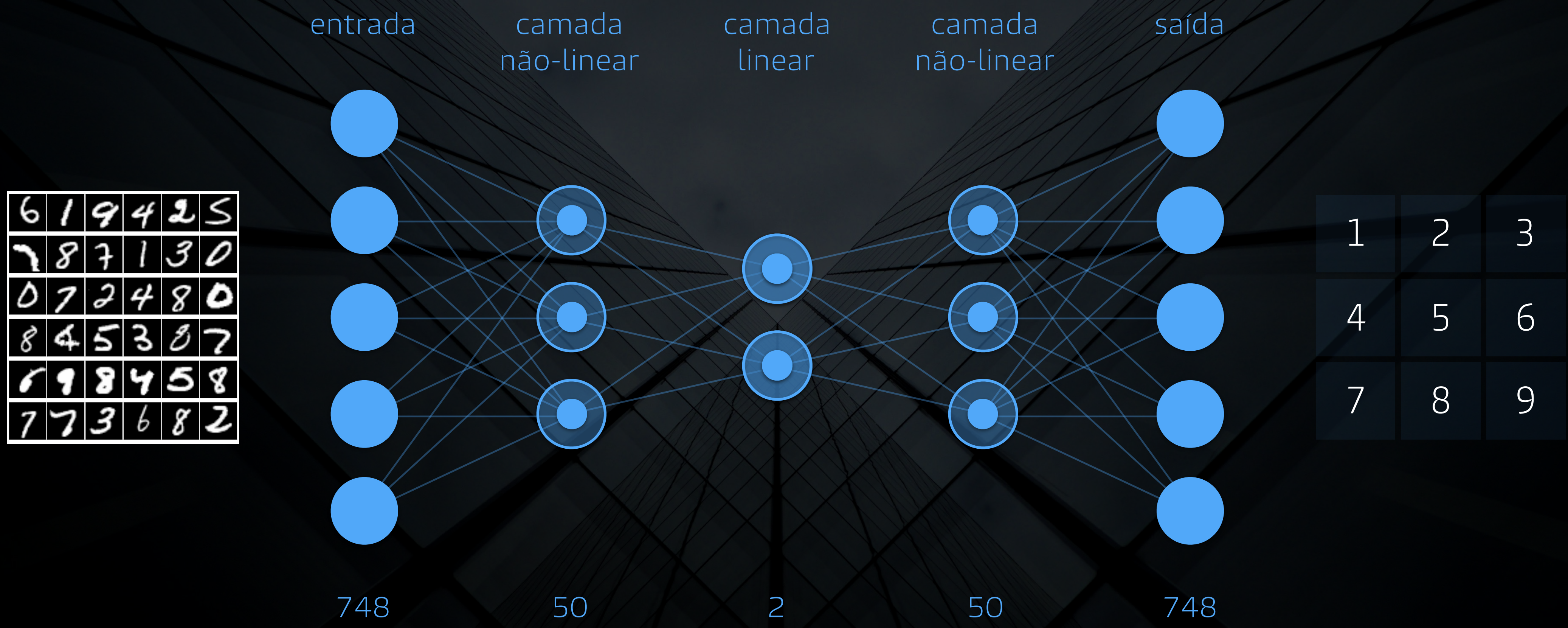

Aprendizado não-supervisionado
Dados → Treino


Aprendizado não-supervisionado
Dados ⟶ Pré-treino

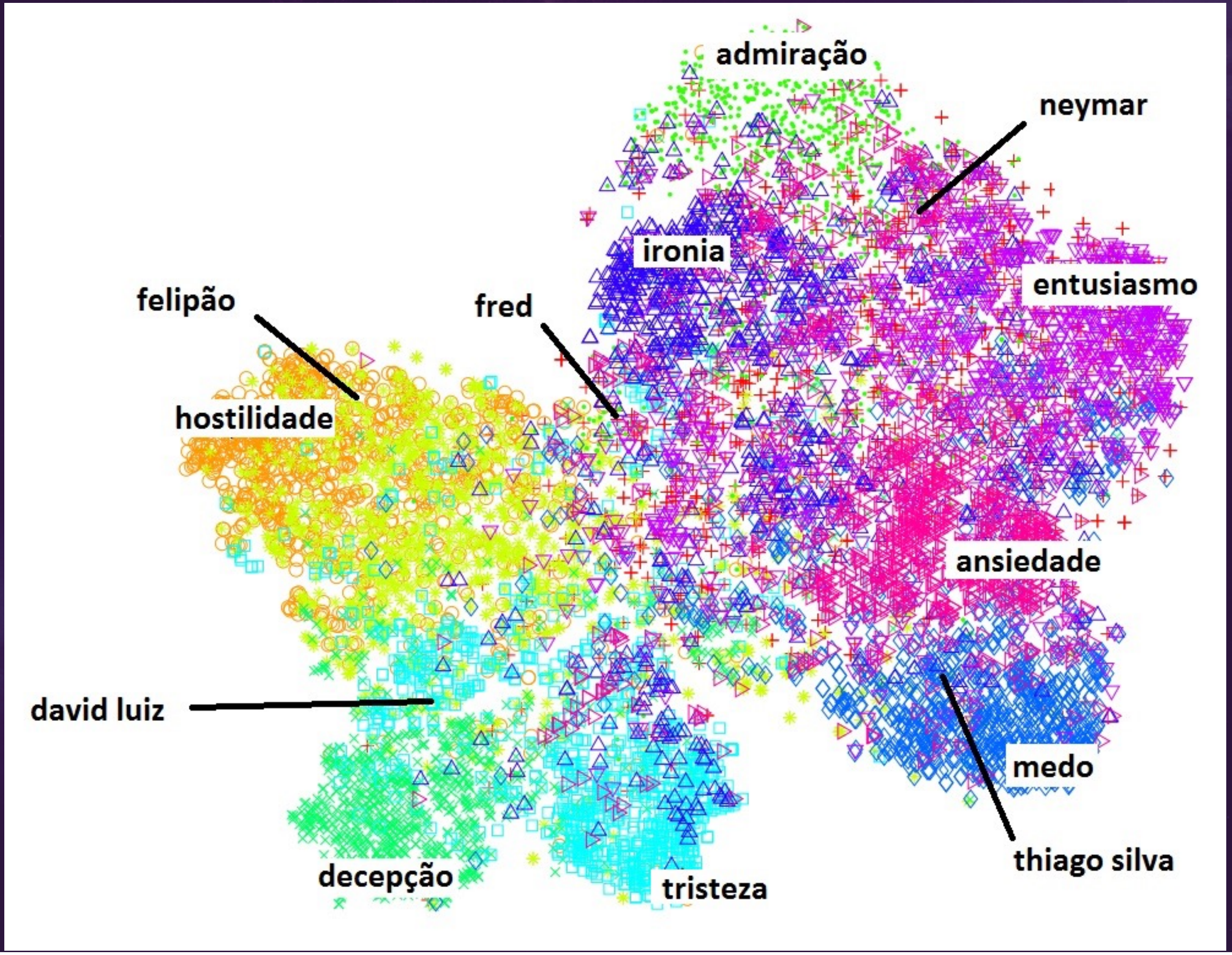

Aprendizado não-supervisionado
Dados → Pré-treino
Dados → Ajuste Fino

Pré-treinamento

Reconhecimento de escrita



Tweets Sobre Jogo Brasil-Alemanha



Redução de Dimensionalidade

Vectoring

Autoencoder (training)

Autoencoder (reduction)

Clustering

Maldição da Dimensionalidade

ID	Idade	Local	Gênero	Mês	Compras	Receita	Ticket Méd.	Filhos	T. na Base
1	45	BH	na	Mai	4	R$ 500,00	R$ 125,00	1	1
2	43	SP	M	Dez	4	R$ 810,00	R$ 202,50	0	0
3	38	SP	na	Fev	3	R$ 675,00	R$ 225,00	2	2
4	36	SP	na	Mai	4	R$ 750,00	R$ 250,00	2	2
5	38	SP	F	Out	2	R$ 843,75	R$ 312,50	0	0
6	41	RJ	M	Abr	4	R$ 450,00	R$ 125,00	3	3
7	37	SP	na	Jun	3	R$ 675,00	R$ 202,50	1	1
8	34	SP	F	Mai	2	R$ 937,50	R$ 225,00	1	1
9	26	BH	na	Mar	2	R$ 450,00	R$ 250,00	1	1
10	34	RJ	F	Ago	2	R$ 450,00	R$ 312,50	1	1
11	31	RJ	M	Jun	3	R$ 562,50	R$ 125,00	3	3
12	28	SP	F	Mar	3	R$ 843,75	R$ 202,50	2	2
13	38	BH	na	Set	2	R$ 450,00	R$ 225,00	0	0
14	37	RJ	na	Dez	3	R$ 540,00	R$ 250,00	1	1
15	27	SP	F	Ago	2	R$ 625,00	R$ 312,50	2	2

Regras que geram o DB

ID	Idade	Local	Gênero	Mês	Compras	Receita	Ticket Méd.	Filhos	T. na Base
1	45								1
2	43								0
3	38								2
4	36								2
5	38								0
6	41								3
7	37								1
8	34								1
9	26								1
10	34								1
11	31								3
12	28								2
13	38								0
14	37								1
15	27	SP	F	Ago	2	R$ 625,00	R$ 312,50	2	2

- **ID:** inteiro de 1 a 90,000
- **Idade:** random de 25 a 45 anos
- **Local:** random entre SP, RJ and BH
- **Gênero:** random [F, M, na], fem., masc. e not-available.
- **Mês:** random Jan a Dec
- **Filhos:** random entre 0 e 3
- **T. na Base:** tempo de cadastro (em meses, entre 1 e 120)
- **Compras:** Idade/10 (inteiro)
- **Ticket Méd:** Valor/Compras

- **Receita**

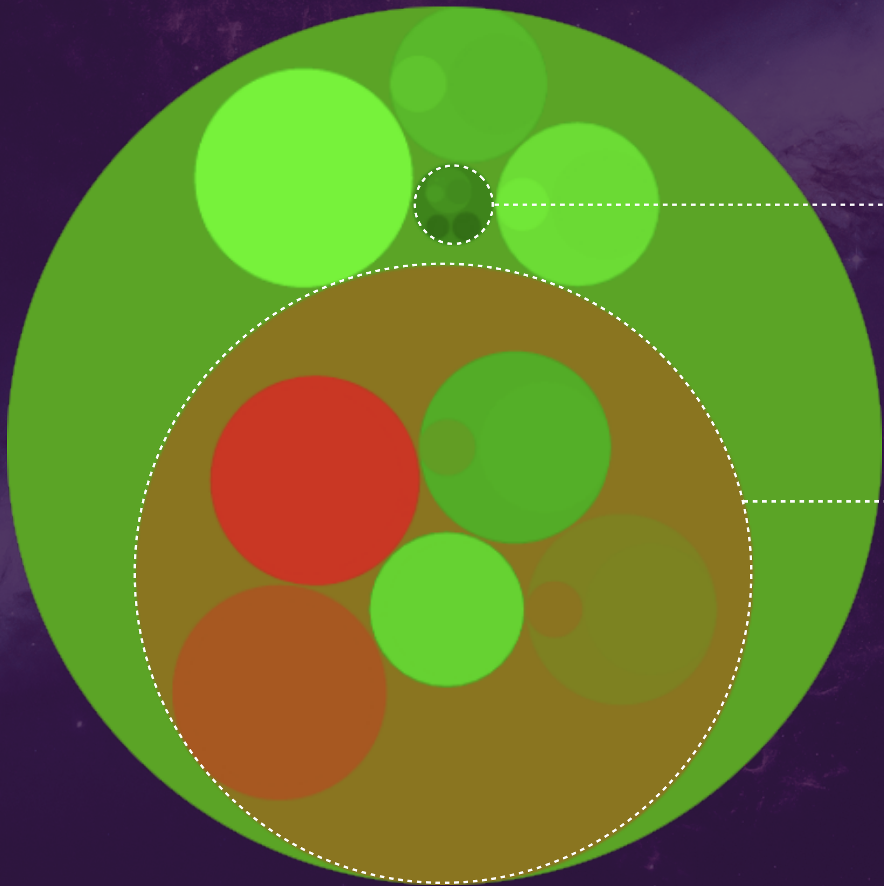
- Começa com R$ 500
- Se Local for SP aumenta 50%
- Se Gênero é F aumenta 25%
- Se Mês é Dez aumenta 20%
- Random entre 90% e 100%
- Nada mais influencia a Receita

Segmentação de Receita

Segmentação
Fatores latentes

- Mulher
- SP
- Dez
- Idade

- R$ 1.125
- R$ 790
- R$ 450



Mulheres, SP, Dez

- Receita: +75.39%
- Ticket M.: R$ 378,60 (+77,81%)

Homens e NA, RJ e BH

- Receita: -11.35%
- Ticket M.: R$ 188,90 (-11,28%)
- Mês: Jun (+3.04%)


precise answers
Qual é o perfil do meu cliente mais rentável?
Mulheres em SP
+75.39% receita
+71.88% ticket médio (R$ 378,60)
qual é o melhor mês para impactar este cliente com uma oferta, de acordo com a curva de venda?
Dezembro
e o perfil menos rentável?
Homens e NA, BH e RJ

Qual é o perfil do meu cliente mais rentável?

What's the profile of my most profitable client?

Mulheres em SP

+75.39% receita

+71.88% ticket médio (R$ 378,60)

Woman in São Paulo
+75.39% revenue
+71.88% avg. ticket
(R$ 378,60)

qual é o melhor mês para impactar este cliente com uma oferta, de acordo com a curva de venda?

what's the best month to impact this client with a new offer, according to his/her sales curve?

Dezembro

December

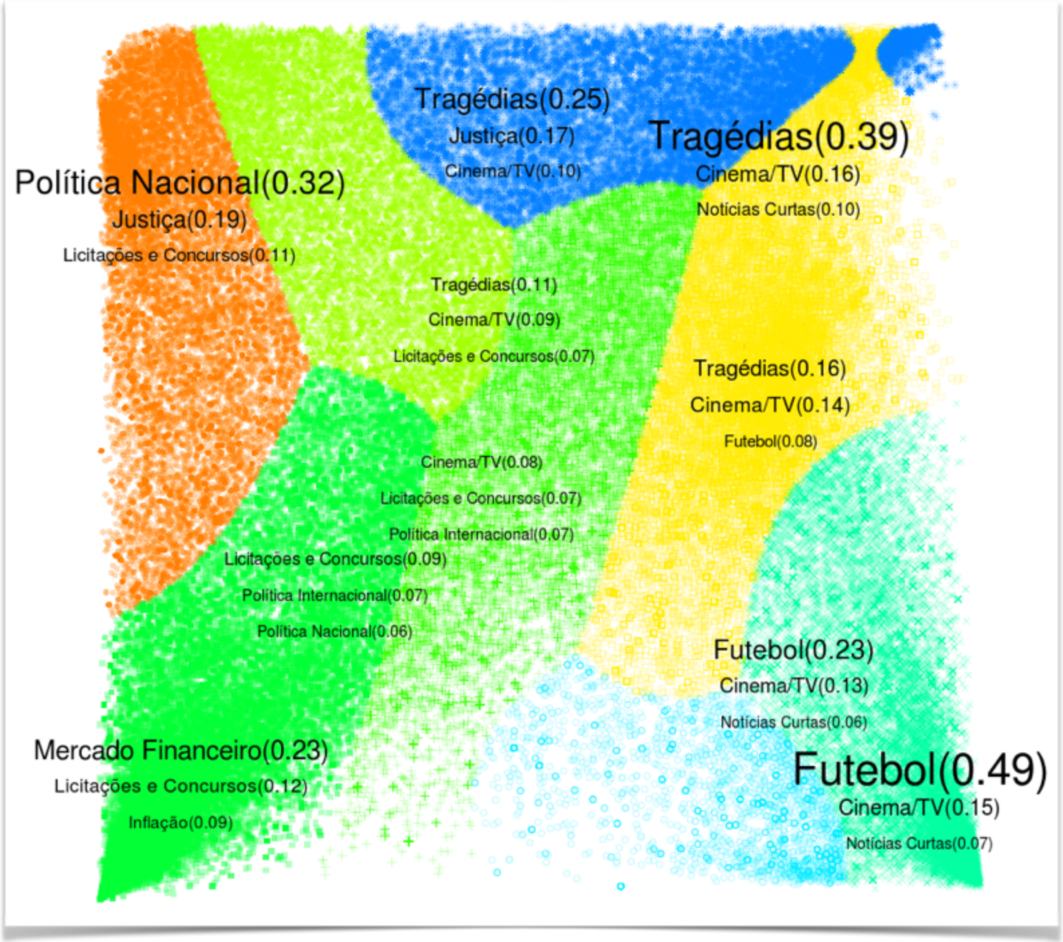
e o perfil menos rentável?

what's the profile of my less profitable client?

Homens e NA, BH e RJ

Man and *"na"*, BH and RJ

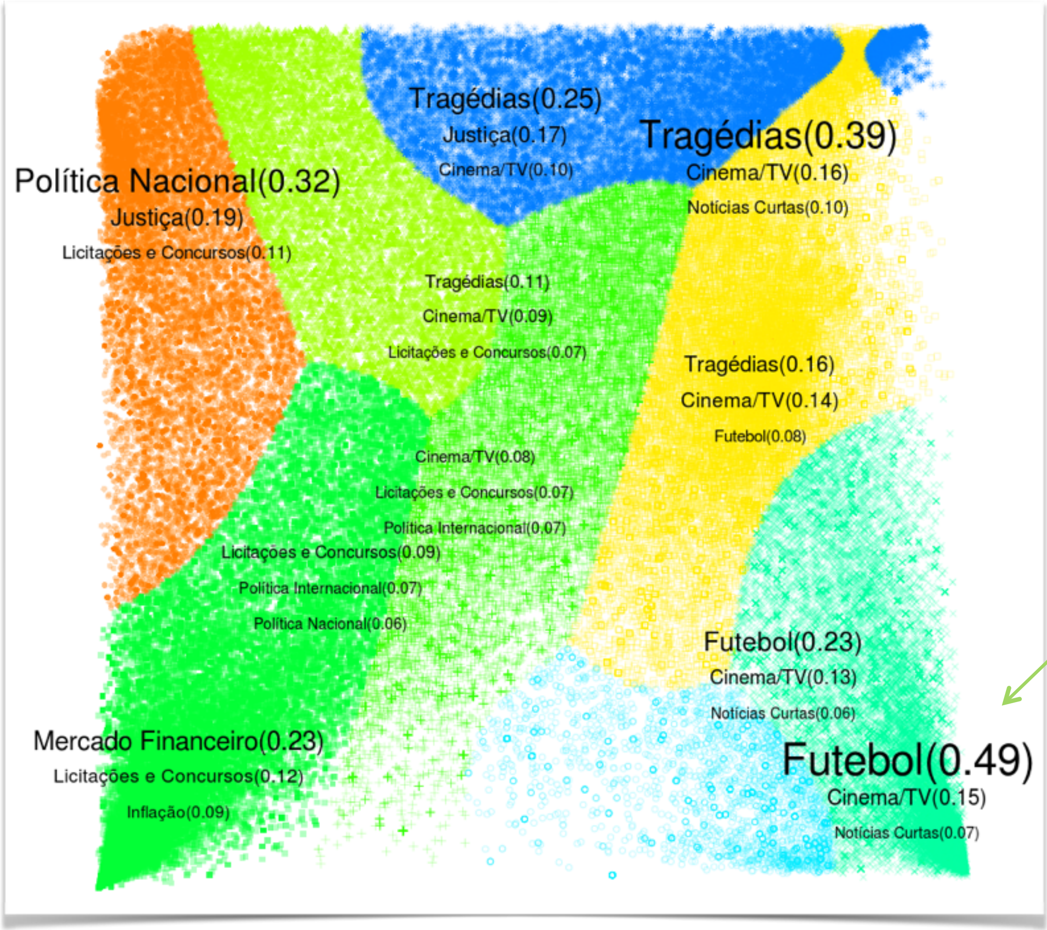
Dimensionality Reduction: Another Example



3.2 million readers

Readers in a cluster have the same preferences

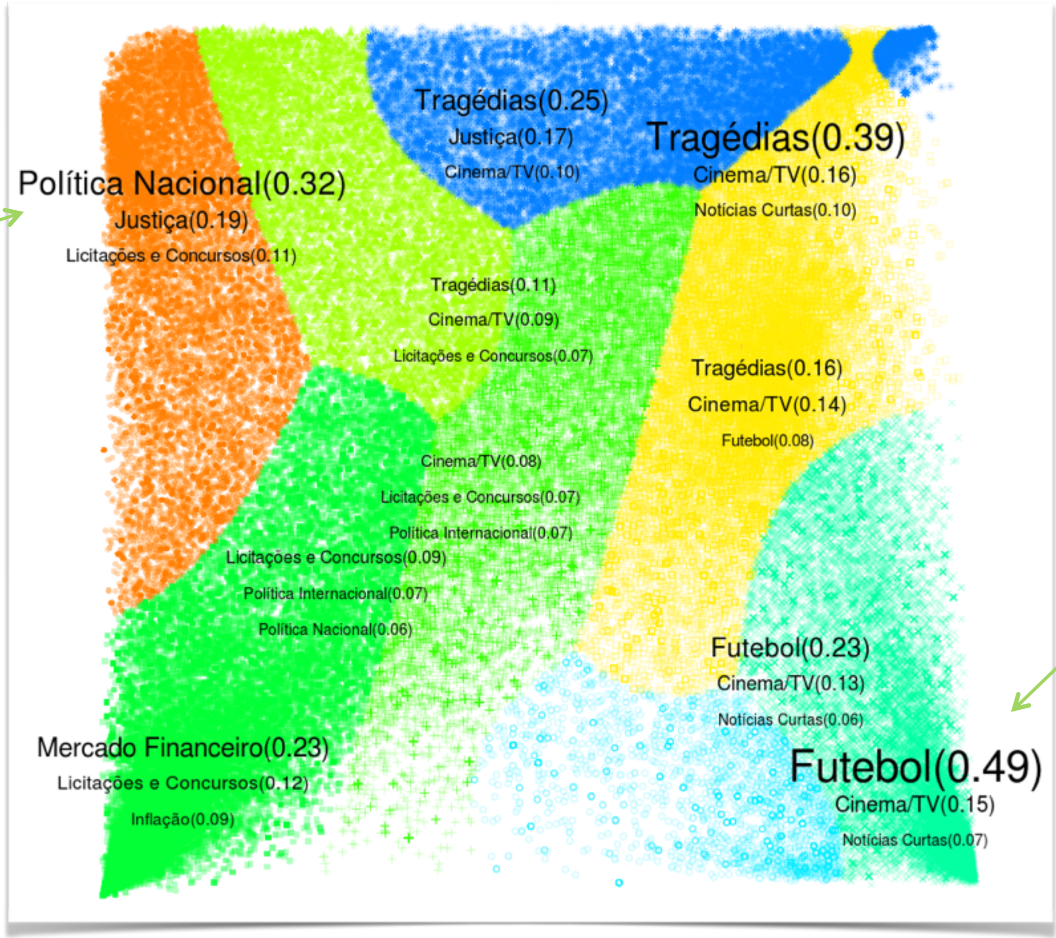
Dimensionality Reduction: Another Example



Soccer 49%
Cinema/TV 15%
Short news 7%
…

Dimensionality Reduction: Another Example

National politics 32%
Justice 19%
Bidding and contests 11%



Soccer 49%
Cinema/TV 15%
Short news 7%
…

kunumi platform

data

Web, SQL, CRM, ERP, Spreadsheets (Excel, Google)

model

Deep Learning algorithms for structured and semi-strucured data analysis

system

pattern recognition systems for data segmentation, trend detection and risk analysis

APIs

integration with enterprise systems (BI, Marketing, HR, Supply Chain, etc)

Some References

Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* 9, 1735–1780 (1997).
Introduces LSTM recurrent networks, a crucial ingredient because they are good at learning long-range dependencies.

Bengio,Y.,Ducharme,R.& Vincent,P. A neural probabilistic language model.In *Proc*. *Advances in Neural Information Processing Systems 13* 932–938 (2001).
Neural language models, which learn to convert a word symbol into a word vector or word embedding composed of learned semantic features to predict the next word in a sequence.

Hinton,G.E., Osindero,S. & Teh,Y.-W. A fast learning algorithm for deep belief nets. *Neural Comp.* **18,** 1527–1554 (2006).
N**ovel way of training very deep neural networks by pre-training one hidden layer at a time using unsupervised learning for restricted Boltzmann machi.**

Some References

Bengio,Y., Lamblin,P., Popovici,D.& Larochelle,H. Greedy layer-wise training of deep networks. In *Advances in Neural Info. Proces Systems 19* 153–160 (2006).
Previous ref. improves performance on test data and generalizes the method to other unsupervised techniques, e.g. auto-encoders.

Glorot, X., Bordes, A. & Bengio. Y. Deep sparse rectifier neural networks. In *14th International Conf. on Artificial Intelligence and Statistics* 315–323 (2011).
Supervised training of very deep neural networks is much faster if the hidden layers are composed of ReLU.

Hinton, G. *et al*. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine* 29, 82–97 (2012).
Joint paper from speech recognition labs, deep learning on phonetic classification for automatic speech recognition, first major industrial application.

Some References

Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks. In *Proc. Advances in Neural Information Processing Systems 27,* 3104–3112 (2014).
M**achine translation results with a recurrent network trained to read a sentence in one language, produce a semantic representation of its meaning, and generate a translation in another language.**

LeCun,Y., Bengio,Y & Hinton ,G. Deep Learning, ***Nature 521,*** **436—444 (2015).**

Some References

Gartner lista 10 tendências tecnológicas de alto impacto para 2016:

http://idgnow.com.br/ti-corporativa/2015/10/26/10-tendencias-tecnologicas-de-alto-impacto-para-2016-segundo-a-gartner/

Yan LeCun - Impressive demonstration of Facebook's AI:

https://www.youtube.com/watch?v=U_Wgc1JOsBk

Conclusions

- Deep learning allows models composed of multiple processing layers to learn representations of data with multiple levels of abstraction
- Improved the state-of-the-art in speech recognition, visual object recognition, object detection, drug discovery, etc.
- Discovers intricate structure in large data sets by using backpropagation to change its internal parameters used to compute the representation in each layer from the representation in previous layer
- Deep convolutional nets presented breakthroughs in processing images, video, speech and audio
- Recurrent nets are recently improving sequential data such as text and speech

UFMG | ICEx
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

Nivio Ziviani
nivio@dcc.ufmg.br

kunumi

nivio@kunumi.com

Kunumi

- **Portfolio of services:**
  - Predictive analysis
  - Business insights
  - Big Data Engineering
- **Solve business problems in areas:**
  - Demand forecast
  - Customer segmentation and targeting
  - Campaign optimisation & effectiveness
  - Inventory planning
  - Social media insights
  - Etc.