"Vejo um tempo em que nós seremos para as máquinas o que os cães são para nós, e eu estou torcendo para as máquinas" (Claude Shannon).

Deep Learning

Paulo Ricardo Lisboa de Almeida

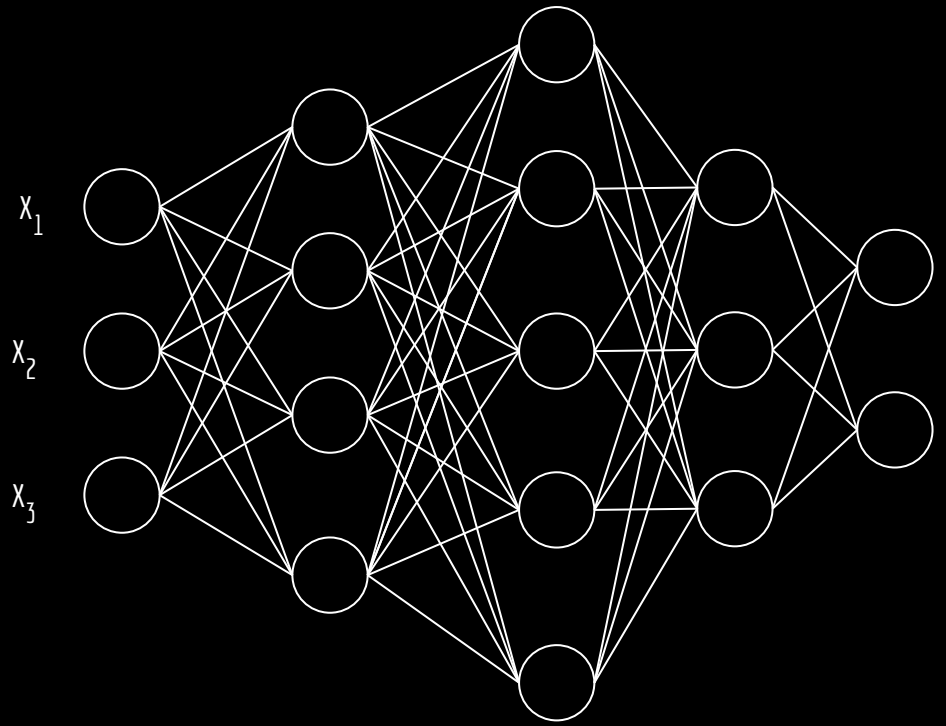
DSBD

UFPR

Deep Learning

Redes neurais com muitas camadas são denominadas profundas – *Deep Neural Networks*.

A área do Aprendizado de Máquina que lida com essas redes é chamada de *Deep Learning* (Aprendizado Profundo).



Deep Learning

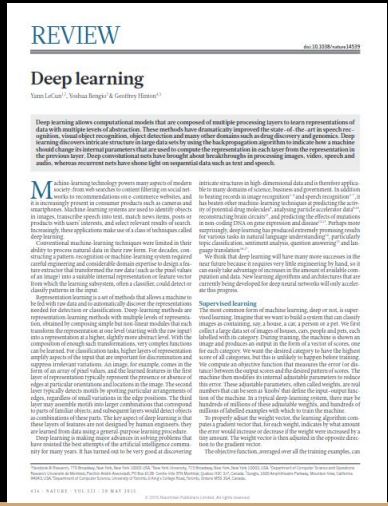
Métodos de aprendizado tradicional possuem limitações com dados crus.

Precisam de especialistas de domínio para transformar o dado cru em algo que pode ser processado.

Extratores de Características.

Métodos de Deep Learning podem “aprender automaticamente” a representação dos dados.

Gerar os extratores de características internamente.



REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations. An image, for example, comes in the form of an array of pixel values, and the learned features in the first layer of representation typically represent the presence or absence of edges at particular orientations and locations in the image. The second layer typically detects motifs by spotting particular arrangements of edges, regardless of small variations in the edge positions. The third layer may assemble motifs into larger combinations that correspond to parts of familiar objects, and subsequent layers would detect objects as combinations of these parts. The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.

Deep learning is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years. It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

We think that deep learning will have many more successes in the near future because it requires very little engineering by hand, so it can easily take advantage of increases in the amount of available computation and data. New learning algorithms and architectures that are currently being developed for deep neural networks will only accelerate this progress.

Supervised learning

The most common form of machine learning, deep or not, is supervised learning. Imagine that we want to build a system that can classify images as containing, say, a house, a car, a person or a pet. We first collect a large data set of images of houses, cars, people and pets, each labelled with its category. During training, the machine is shown an image and produces an output in the form of a vector of scores, one for each category. We want the desired category to have the highest score of all categories, but this is unlikely to happen before training. We compute an objective function that measures the error (or distance) between the output scores and the desired pattern of scores. The machine then modifies its internal adjustable parameters to reduce this error. These adjustable parameters, often called weights, are real numbers that can be seen as 'knobs' that define the input–output function of the machine. In a typical deep-learning system, there may be hundreds of millions of these adjustable weights, and hundreds of millions of labelled examples with which to train the machine.

To properly adjust the weight vector, the learning algorithm computes a gradient vector that, for each weight, indicates by what amount the error would increase or decrease if the weight were increased by a tiny amount. The weight vector is then adjusted in the opposite direction to the gradient vector.

The objective function, averaged over all the training examples, can

[1]Facebook AI Research, 770 Broadway, New York, New York 10003, USA. [2]New York University, 715 Broadway, New York, New York 10003, USA. [3]Department of Computer Science and Operations Research, Université de Montréal, Montréal, Québec H3C 3J7, Canada. [4]Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA. [5]Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, Ontario M5S 3G4, Canada.

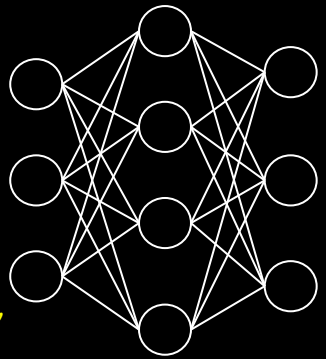
LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553: 436-444. 2015.

Machine Learning Convencional

Extrair características.
Ex.: histogramas de cores.

0	503
1	527
2	326
...	...
254	981
255	1038

Alimentar um Classificador (MLP, SVM, KNN, ...).



Resposta.

Carro.

Não Existe Almoço Grátis

No Machine Learning “Convencional”.

- Em teoria podemos criar modelos tão bons quanto os de Deep Learning.
- Os modelos geralmente são computacionalmente baratos.
- Temos controle sobre o pipeline.
    - Mais fácil entender o que está acontecendo, e o motivo de determinado modelo ter chego a determinada conclusão.

Não Existe Almoço Grátis

No Machine Learning “Convencional”.

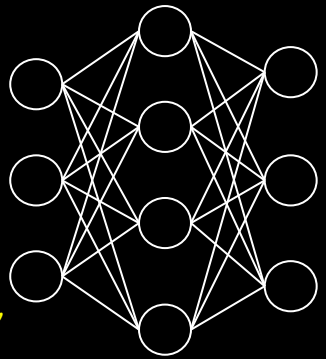
+ Em teoria podemos criar modelos tão bons quanto os de Deep Learning.
+ Os modelos geralmente são computacionalmente baratos.
+ Temos controle sobre o pipeline.
    + Mais fácil entender o que está acontecendo, e o motivo de determinado modelo ter chego a determinada conclusão.
- Muitas vezes é difícil fazer a engenharia de características.
    - Como definir o melhor extrator para o problema em questão.
    - Dependente do problema.
- Complexo.
    - Fazer a engenharia de características e criar um bom pipeline exige conhecimento profundo de ML e do problema em questão.

Machine Learning Convencional

Extrair características.
Ex.: histogramas de cores.

0	503
1	527
2	326
...	...
254	981
255	1038

Alimentar um Classificador (MLP, SVM, KNN, ...).



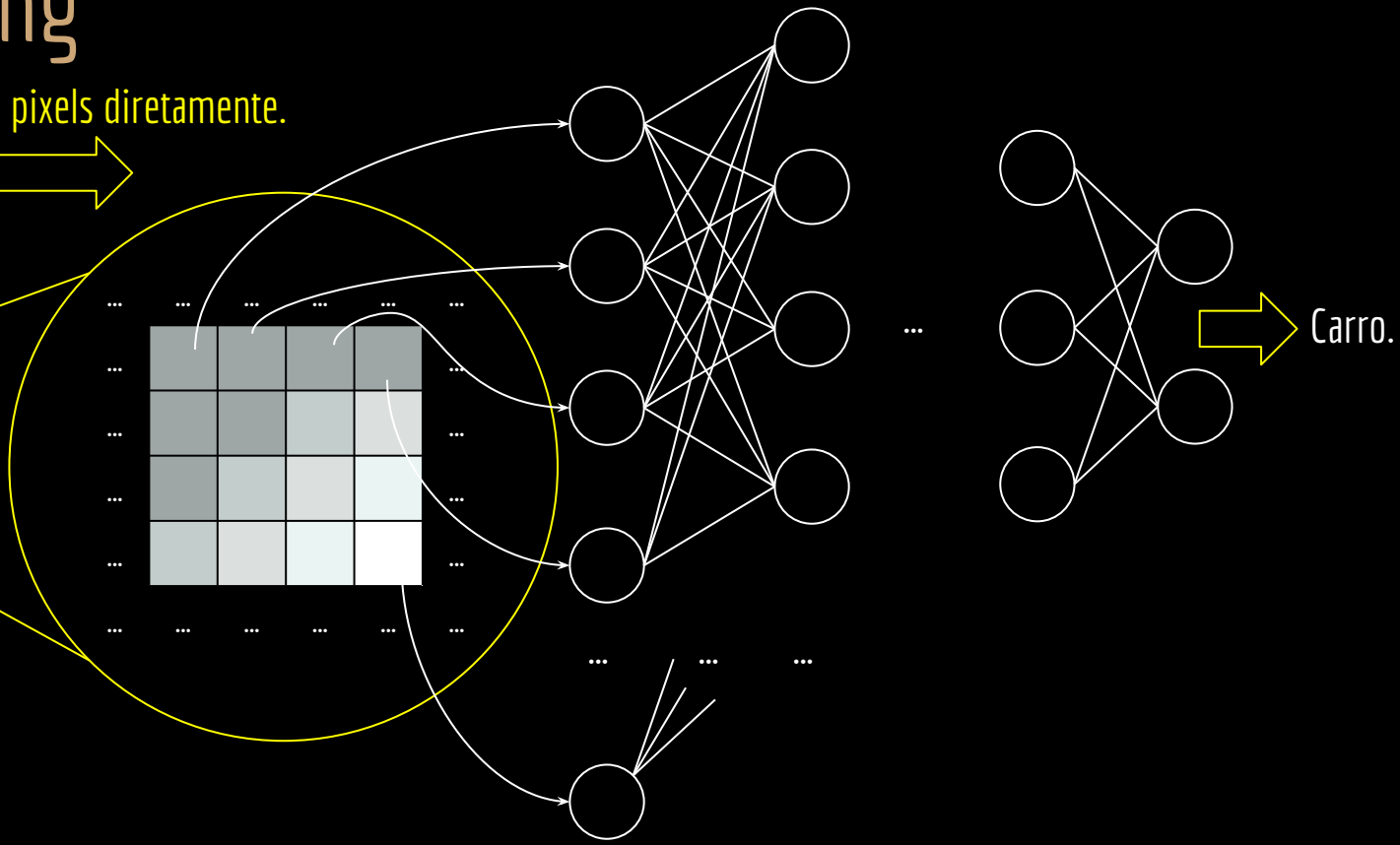
Resposta.

Carro.

Usamos um histograma de cores. Mas por que não Momentos de Hu? Características de Textura? Histograma de Gradientes Orientados? ...

Deep Learning

Podemos alimentar os pixels diretamente.



Deep Learning

Esperamos que as camadas ocultas aprendam a “extrair características”.

Carro.

Deep Learning

Como com Deep Learning é necessário que a rede ainda aprenda uma representação interna dos dados, comumente precisamos:

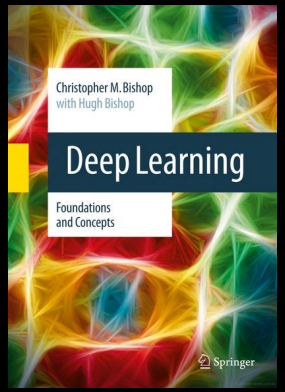
- Mais dados de treinamento quando comparado a modelos tradicionais.
- Maior poder computacional.

Número de Parâmetros

Rede	Parâmetros	Memória (float 32 bits)
MobileNetV3	5 Milhões	0,02 GB
ResNet152	6 Milhões	0,22 GB
vit_l_32	306 Milhões	1,14 GB
Llama 3 8B	8 Bilhões	27 GB
Llama 3 70B	70 Bilhões	261 GB
GPT-3	175 Bilhões	651 GB

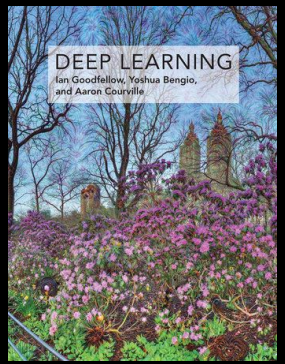
Referências

Bishop, C. M., Bishop, H. Deep Learning: Foundations and Concepts. 2023.

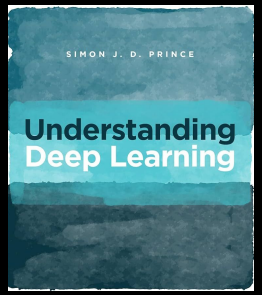


LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553: 436-444. 2015.

Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. 2016.



Prince, S. J. Understanding Deep Learning. 2023.



Licença