The mathematical building blocks of neural networks

Rodrigo Gonzalez, PhD

Hello!

I am Rodrigo Gonzalez, PhD

You can find me at rodralez@gmail.com

1.
What is deep learning?

Artificial intelligence, machine learning, and deep learning

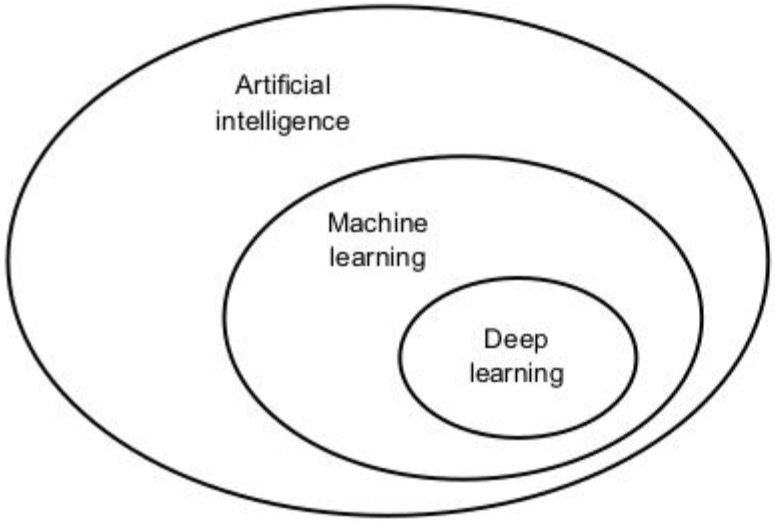


Figure 1.1 Artificial intelligence, machine learning, and deep learning

Machine learning

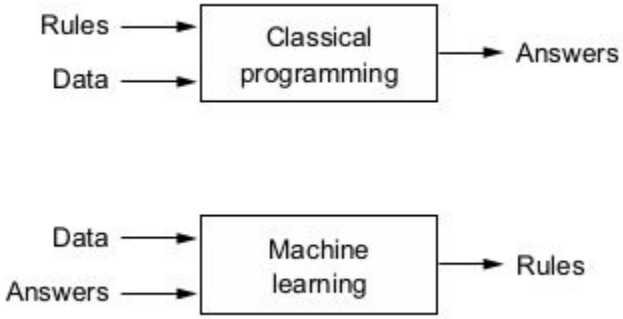


Figure 1.2 Machine learning: a new programming paradigm

Deep learning

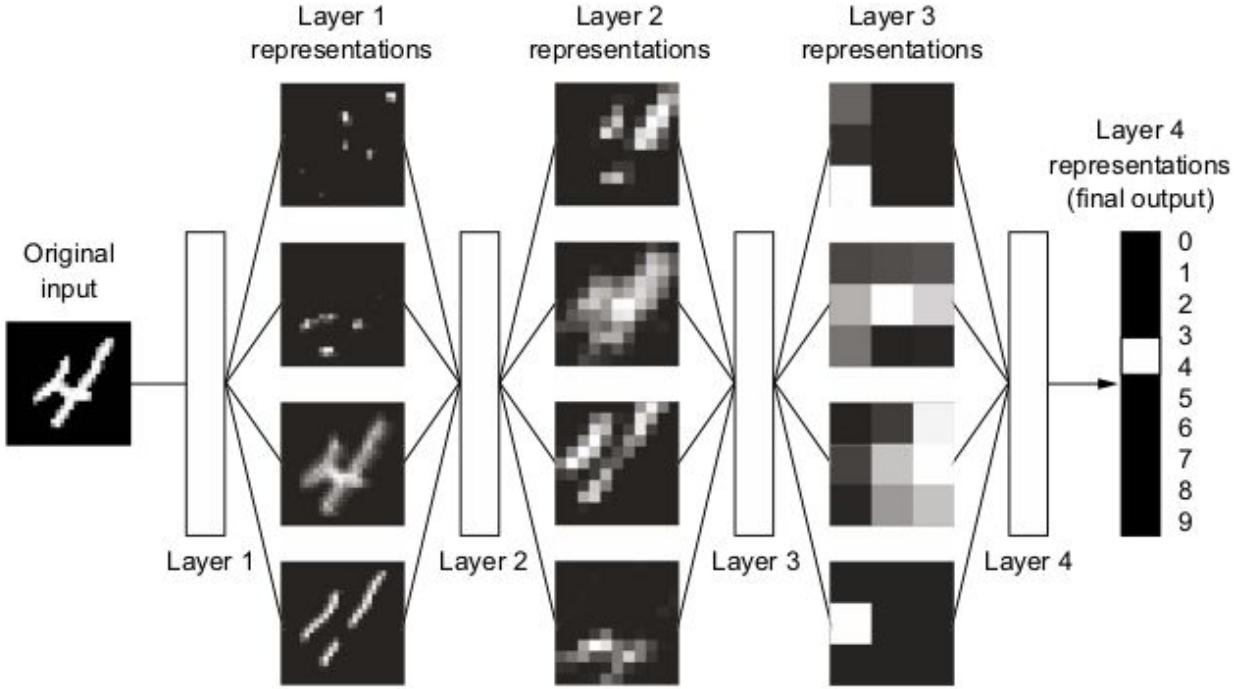


Figure 1.6 Data representations learned by a digit-classification model

How deep learning works

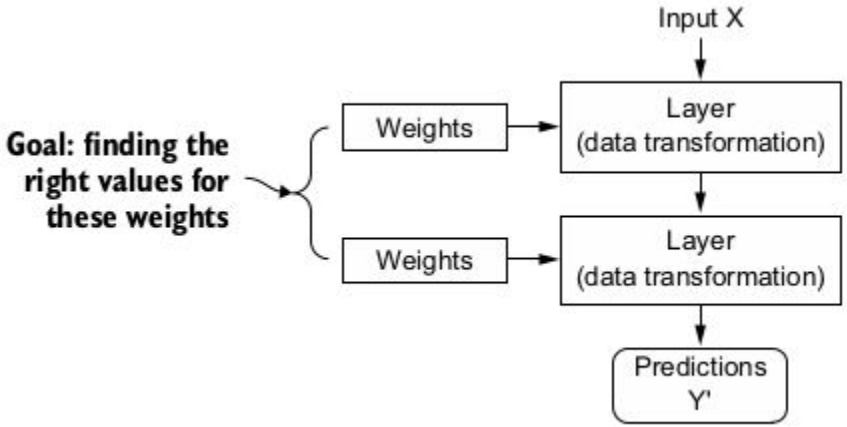
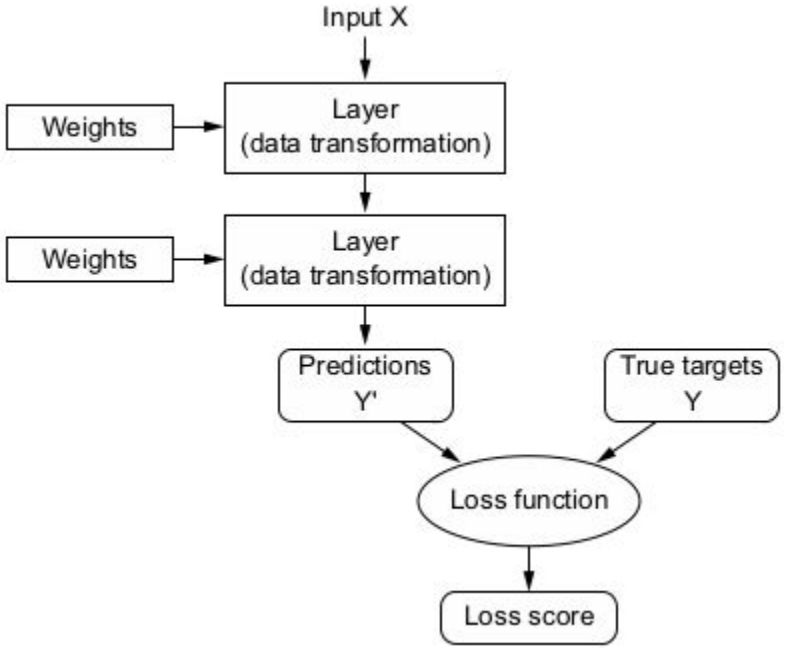
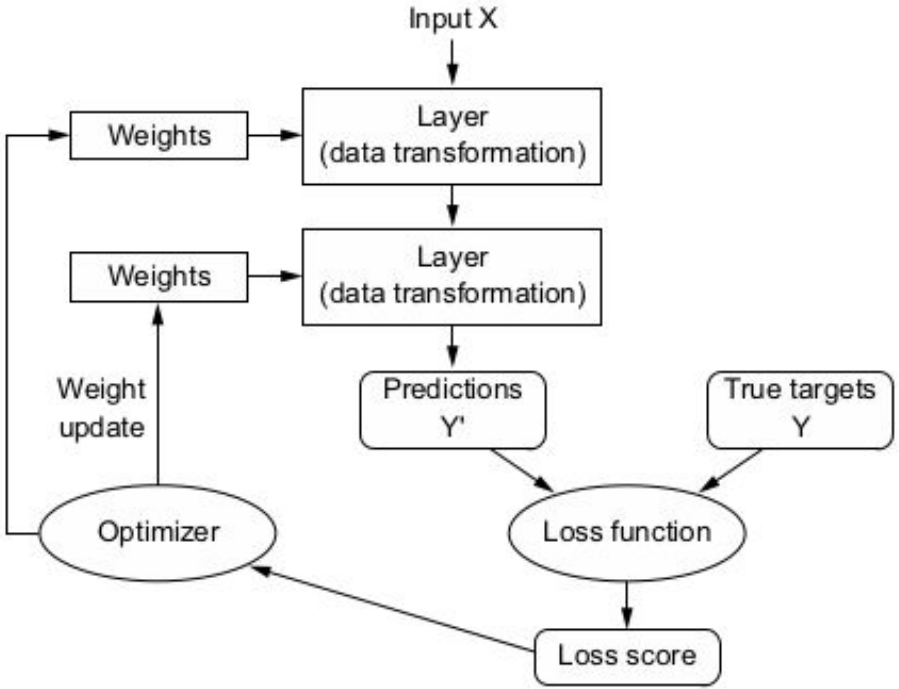


Figure 1.7 A neural network is parameterized by its weights.

How deep learning works



How deep learning works



What deep learning has achieved so far

- Near-human-level image classification
- Near-human-level speech transcription
- Near-human-level handwriting transcription
- Dramatically improved machine translation
- Dramatically improved text-to-speech conversion
- Digital assistants such as Google Assistant and Amazon Alexa
- Near-human-level autonomous driving
- Improved ad targeting, as used by Google, Baidu, or Bing
- Improved search results on the web
- Ability to answer natural language questions
- Superhuman Go playing

2.

A first look at a neural network

The “Hello World” of deep learning

- MNIST dataset, grayscale images of handwritten digits (28 × 28 pixels)
- 60,000 training images, plus 10,000 test images
- Goal: to classify images of handwritten digits into their 10 categories (0 through 9).
- In machine learning,
  - A category in a classification problem is called a **class**.
  - Data points are called **samples**.
  - The class associated with a specific sample is called a **label**.

Workflow

1. First, we'll feed the neural network the training data, train_images and train_labels.
2. The network will then learn to associate images and labels.
3. Finally, we'll ask the network to produce predictions for test_images, and we'll verify whether these predictions match the labels from test_labels.

Layer in neural network

- **Layers** extract representations out of the data fed into them—hopefully, representations that are more meaningful for the problem at hand.
- Most of deep learning consists of chaining together simple layers that will implement a form of progressive *data distillation*.
- A deep learning model is like a sieve for data processing, made of a succession of increasingly refined data filters—the layers.

Compilation step

- An **optimizer**: The mechanism through which the model will update itself based on the training data it sees, so as to improve its performance.
- A **loss function**: How the model will be able to measure its performance on the training data, and thus how it will be able to steer itself in the right direction.
- **Metrics** to monitor during training and testing: Here, we'll only care about accuracy (the fraction of the images that were correctly classified).

3.

Data representations for neural networks

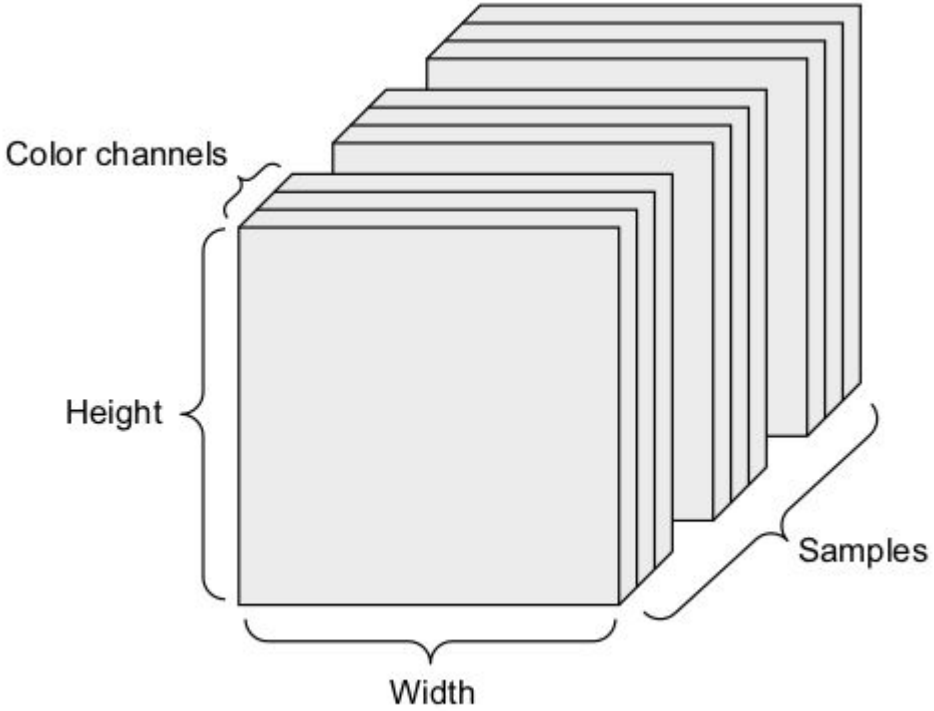
Data representations for neural networks

- Scalars (rank-0 tensors)
- Vectors (rank-1 tensors)
- Matrices (rank-2 tensors)
- Rank-3 and higher-rank tensors

Key attributes

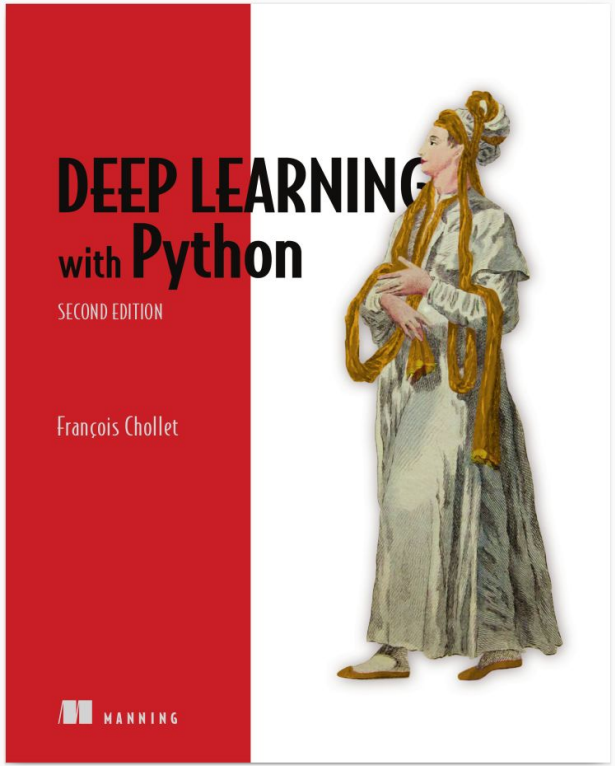
- Number of axes or rank (`ndim`)
- Shape (`shape`) tuple of integers that describes how many dimensions the tensor has along each axis.
- Data type (`dtype`), float16, float32, float64, uint8.

Image data



4.
The book

Deep Learning with Python, 2nd Ed. by Francois Chollet



- Chapter 2

Thanks!

Any questions?

You can find me at:

rodralez@gmail.com