Deep Learning

Reference book

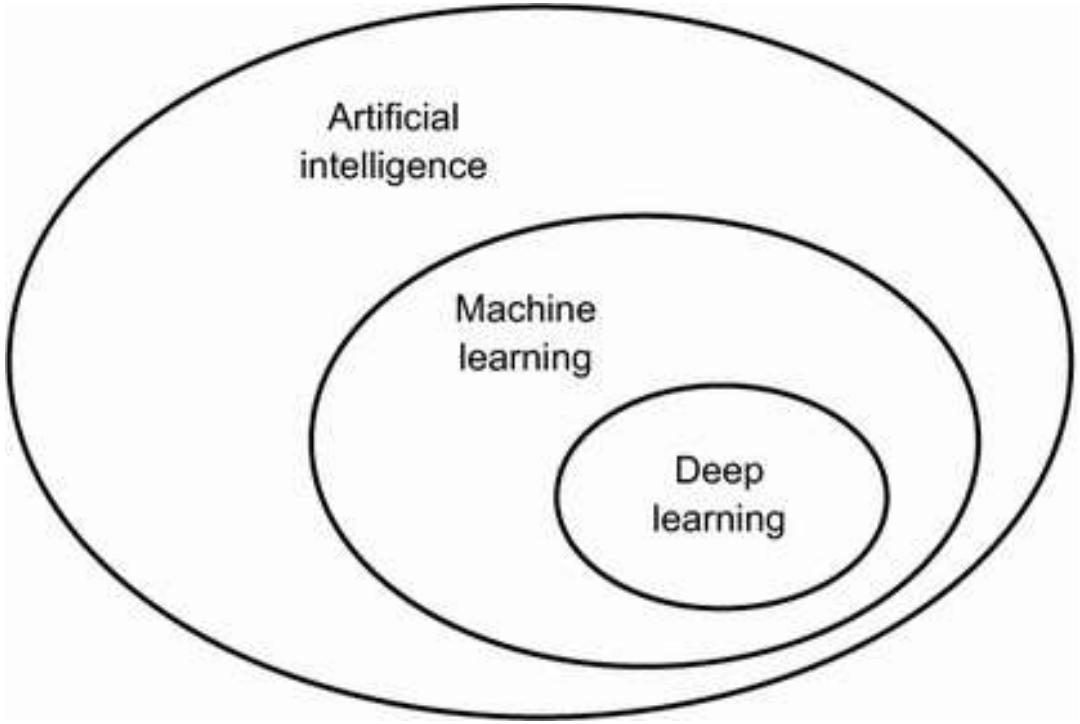
- *Deep Learning with Python*, Second Edition by François Chollet

Introduction to Deep Learning

Artificial Intelligence, Machine Learning, and Deep Learning

- First, we need to define clearly what we’re talking about when we mention AI.
- What are artificial intelligence, machine learning, and deep learning?
- How do they relate to each other?

Relating AI, ML, and DL



Artificial intelligence, machine learning, and deep learning

Artificial Intelligence (AI)

- AI was born in the 1950s.
- "artificial intelligence" crystallized as a field of research in 1956, when John McCarthy, then a young Assistant Professor of Mathematics at Dartmouth College, organized a summer workshop under the following proposal:
  - The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves...

Artificial Intelligence (AI)

- The effort to automate intellectual tasks normally performed by humans.
- A general field that encompasses machine learning and deep learning
- But that also includes many more approaches that may not involve any learning.

Symbolic AI

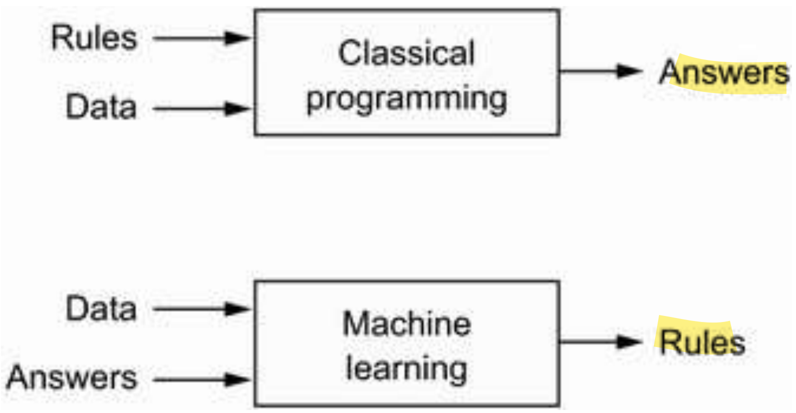
- Used mostly hardcoded rules crafted by programmers – Chess game
- It reached its peak popularity during the expert systems boom of the 1980s.

Symbolic AI to Machine Learning

- Symbolic AI proved suitable to solve well-defined, logical problems (e.g. chess).
- It turned out to be intractable to figure out explicit rules for solving more complex problems (e.g. image classification, speech recognition, or natural language translation)
- A new approach arose to take symbolic AI's place: machine learning (ML).

Symbolic AI to Machine Learning

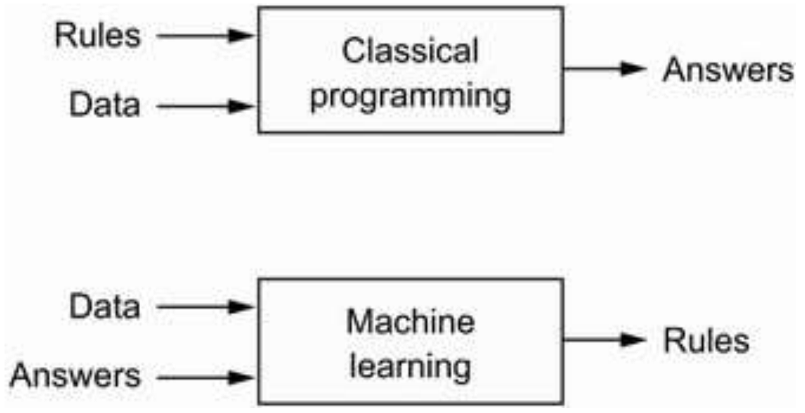
- Classical programming is to have a human programmer write down *rules*—a computer program—to be followed to turn input data into appropriate answers.
- Machine Learning turns this around.
- The machine looks at the input data and the corresponding answers, and figures out what the rules should be.



Machine learning:
a new programming paradigm

Machine Learning (ML)

- A machine learning system is *trained* rather than explicitly programmed.
- It's presented with many examples relevant to a task, and it finds statistical structure in these examples that eventually allows the system to come up with rules for automating the task.



Machine learning:
a new programming paradigm

Machine Learning (ML)

- Machine learning only started to flourish in the 1990s.
- Subfield of AI, a trend driven by the availability of faster hardware and larger datasets.
- ML is related to mathematical statistics, but it differs from statistics in several important ways.
- Machine learning is a very hands-on field driven by empirical findings and deeply reliant on advances in software and hardware.

Learning rules and representations from data

- To do machine learning, we need three things:
  - Input data points
  - Examples of the expected output
  - A way to measure whether the algorithm is doing a good job

Learning rules and representations from data

- Input data points
  - if the task is speech recognition, these data points could be sound files of people speaking.
  - If the task is image tagging, they could be pictures.

Learning rules and representations from data

- Examples of the expected output
  - In a speech-recognition task, these could be human-generated transcripts of sound files.
  - In an image task, expected outputs could be tags such as “dog,” “cat,” and so on.

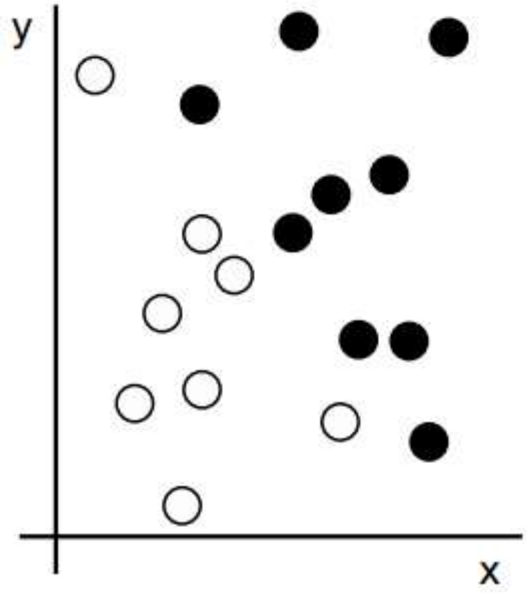
Learning rules and representations from data

- A way to measure whether the algorithm is doing a good job
  - This is necessary in order to determine the distance between the algorithm’s current output and its expected output.
  - The measurement is used as a feedback signal to adjust the way the algorithm works.
  - This adjustment step is what we call *learning*.

Learning rules and representations from data

- A machine learning model transforms its input data into meaningful outputs, a process that is "learned" from exposure to known examples of inputs and outputs.
- Therefore, the central problem in machine learning and deep learning is to *meaningfully transform data*:
  - to learn useful *representations* of the input data at hand—representations that get us closer to the expected output.

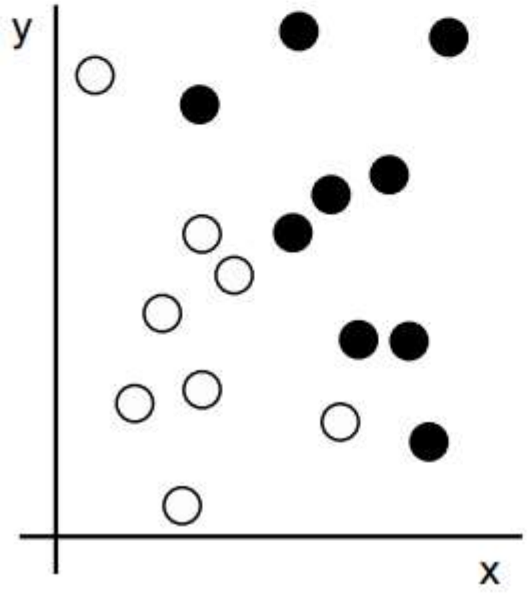
Learning rules and representations from data



Some sample data

- The inputs are the coordinates of our points.
- The expected outputs are the colors of our points.
- A way to measure whether our algorithm is doing a good job could be, for instance, the percentage of points that are being correctly classified.

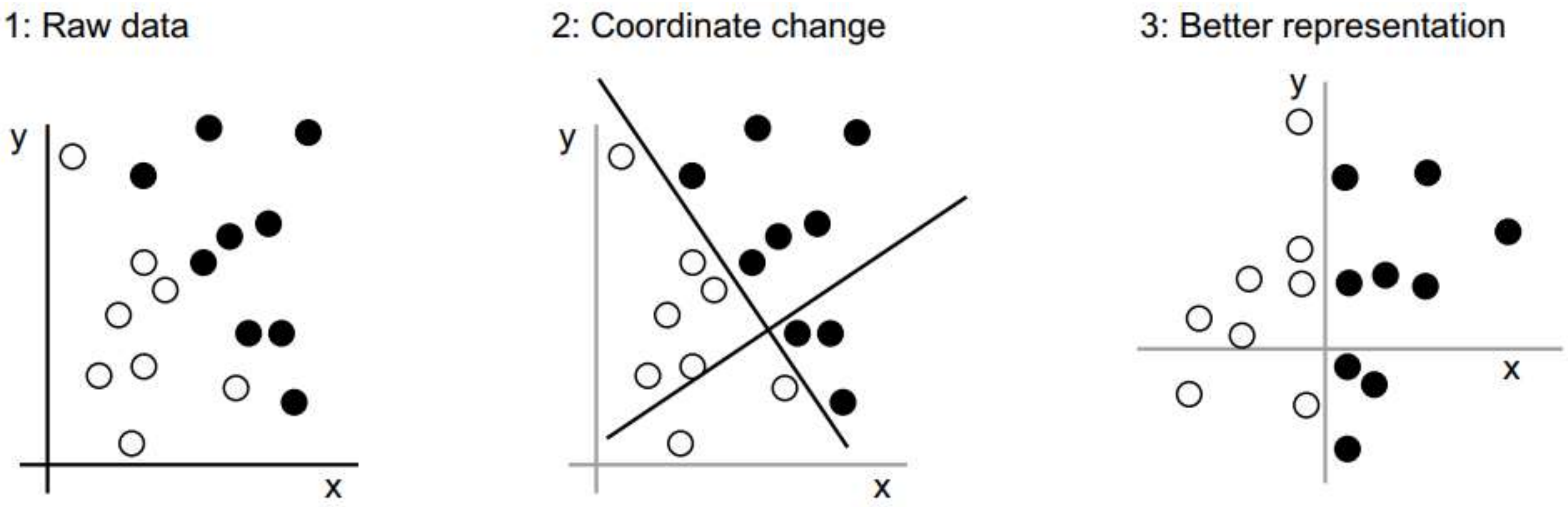
Learning rules and representations from data



Some sample data

- What we need here is a new representation of our data that cleanly separates the white points from the black points.
- One transformation we could use, among many other possibilities, would be a coordinate change .

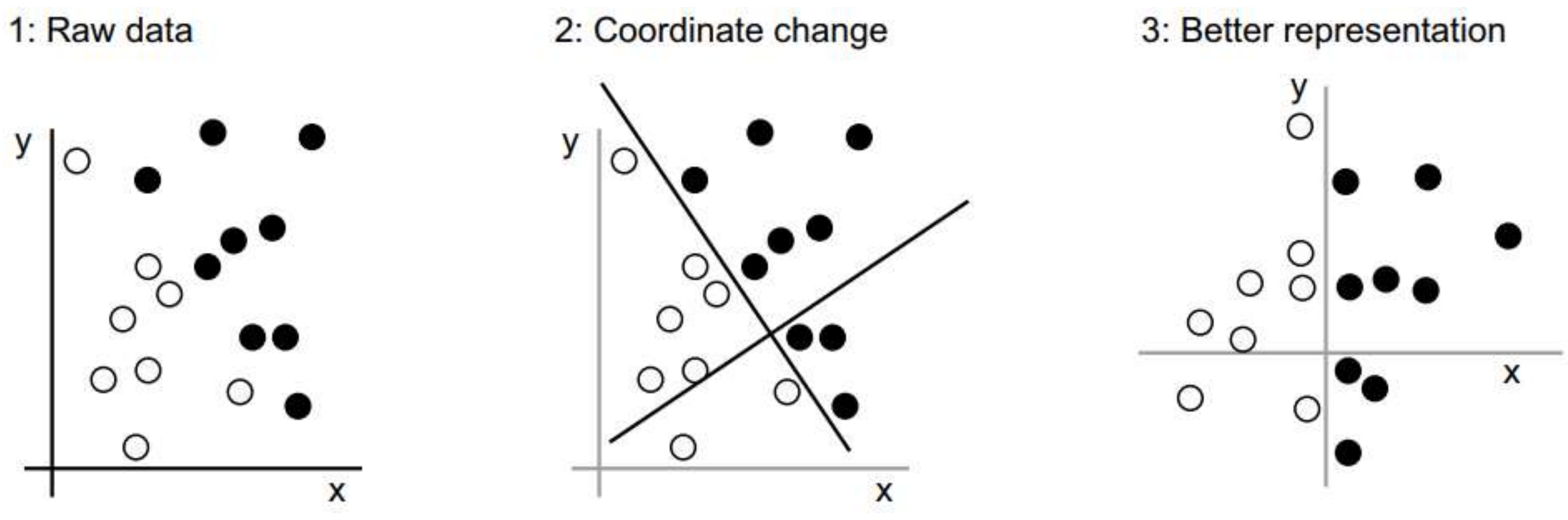
Learning rules and representations from data



Coordinate change

- In this new coordinate system, the coordinates of our points can be said to be a new representation of our data.

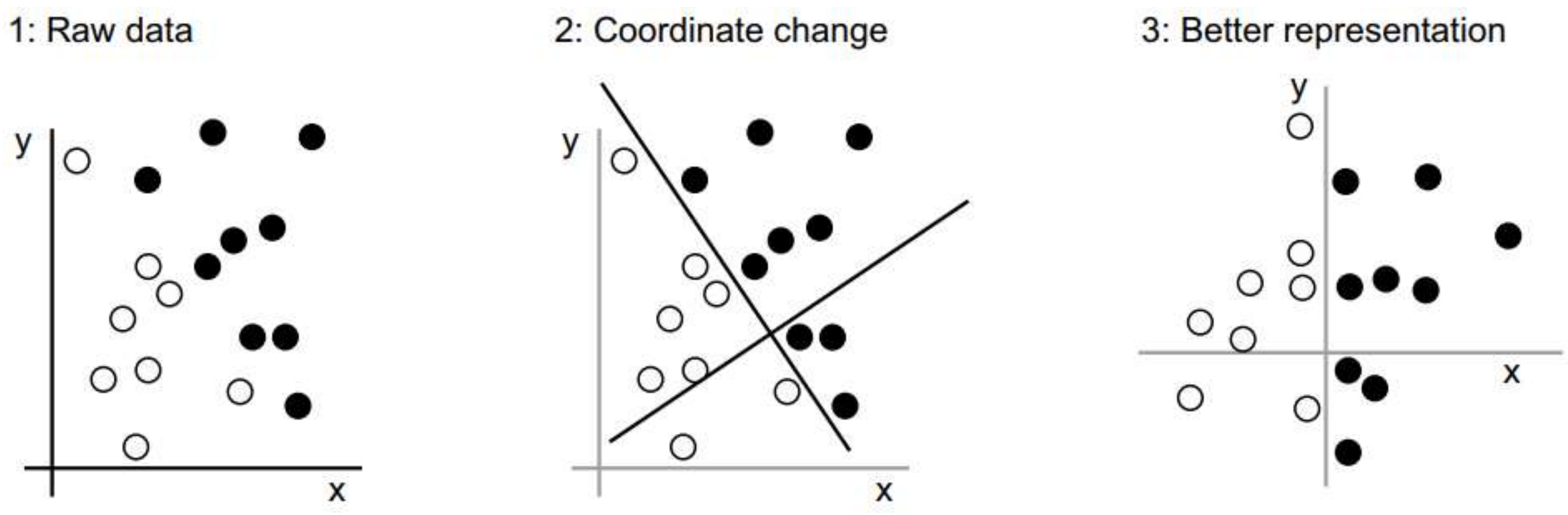
Learning rules and representations from data



Coordinate change

- With this representation, the black/white classification problem can be expressed as a simple rule: "Black points are such that $x > 0$," or "White points are such that $x < 0$."

Learning rules and representations from data



Coordinate change

- This new representation, combined with this simple rule, neatly solves the classification problem.

Learning rules and representations from data

- In this case we defined the coordinate change by hand: we used our human intelligence to come up with our own appropriate representation of the data.
- Could we automate it?
- What if we tried systematically searching for different sets of automatically generated representations of the data and rules based on them, identifying good ones by using as feedback the percentage of digits being correctly classified in some development dataset?

Learning rules and representations from data

- We would then be doing machine learning.
- *Learning*, in the context of machine learning, describes an automatic search process for data transformations that produce useful representations of some data, guided by some feedback signal—representations that are amenable to simpler rules solving the task at hand.

Learning rules and representations from data

- Machine learning algorithms aren’t usually creative in finding these transformations
- They’re merely searching through a predefined set of operations, called a *hypothesis space*.
- For instance, the space of all possible coordinate changes would be our hypothesis space in the 2D coordinates classification example.

Learning rules and representations from data

- So that’s what machine learning is, concisely:
  - Searching for useful representations and rules over some input data, within a predefined space of possibilities, using guidance from a feedback signal.
- This simple idea allows for solving a remarkably broad range of intellectual tasks, from speech recognition to autonomous driving.

Deep Learning (DL)

- A subfield of ML.
- A new take on learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations.

Deep Learning (DL)

- The “deep” in “deep learning” isn’t a reference to any kind of deeper understanding achieved by the approach;
- rather, it stands for this idea of successive layers of representations.
- How many layers contribute to a model of the data is called the *depth* of the model.
- Other appropriate names for the field could have been *layered representations learning* or *hierarchical representations learning*.

Deep Learning (DL)

- Modern deep learning often involves tens or even hundreds of successive layers of representations, and they're all learned automatically from exposure to training data.
- Meanwhile, other approaches to machine learning tends to focus on learning only one or two layers of representations of the data called shallow learning.

Deep Learning (DL)

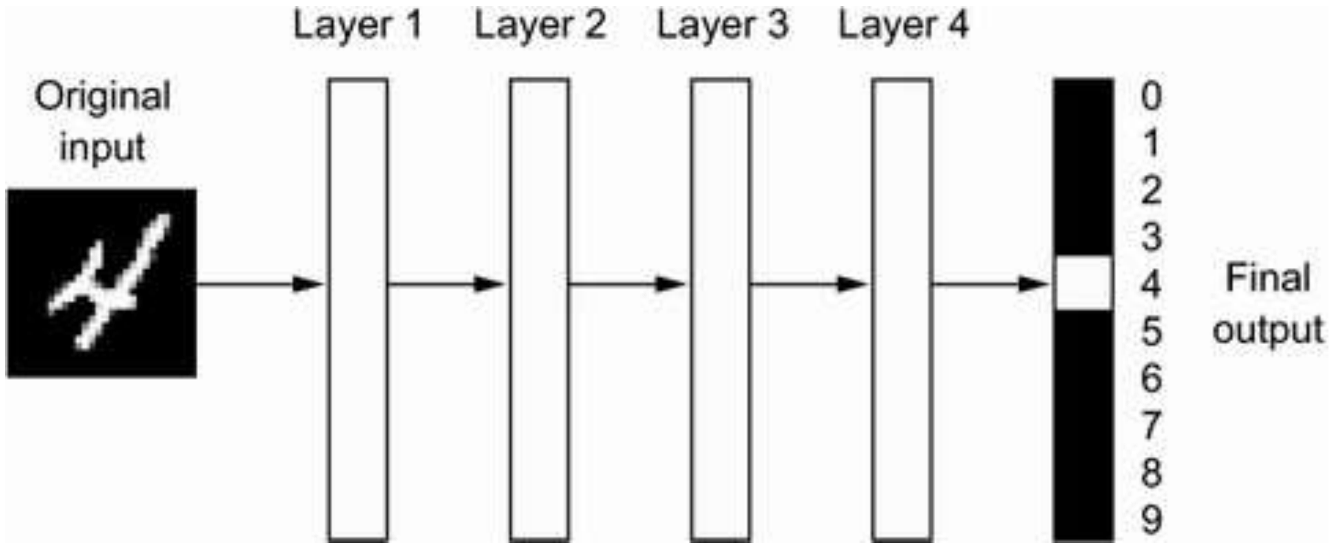
- In deep learning, these layered representations are learned via models called *neural networks*, structured in literal layers stacked on top of each other.
- The term “neural network” refers to neurobiology, but although some of the central concepts in deep learning were developed in part by drawing inspiration from our understanding of the brain (in particular, the visual cortex), deep learning models are not models of the brain.
- There’s no evidence that the brain implements anything like the learning mechanisms used in modern deep learning models.

Deep Learning (DL)

- For our purposes, deep learning is a mathematical framework for learning representations from data.
- What do the representations learned by a deep learning algorithm look like?

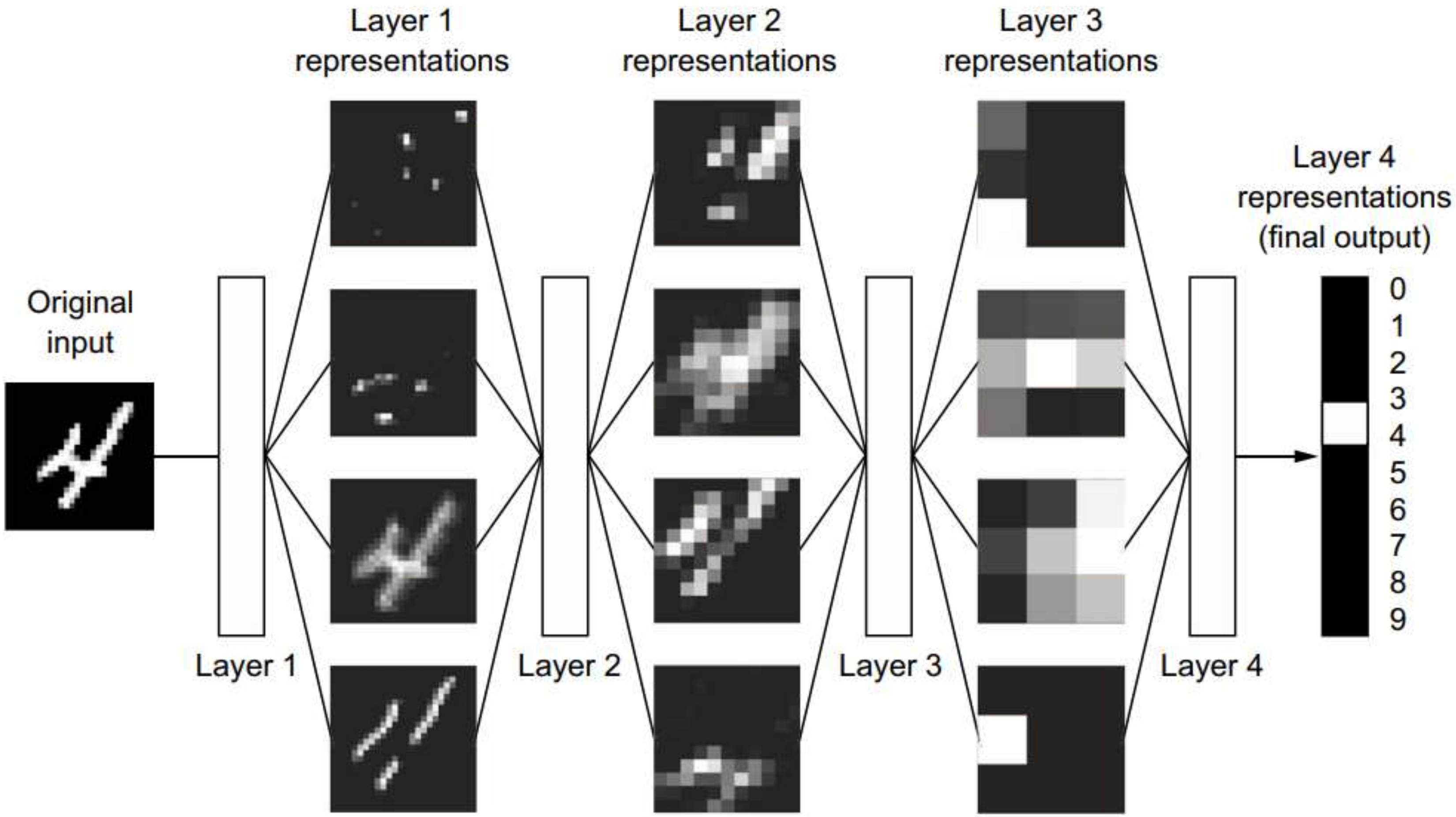
Deep Learning Network

- Let's examine how a network several layers deep transforms an image of a digit in order to recognize what digit it is.



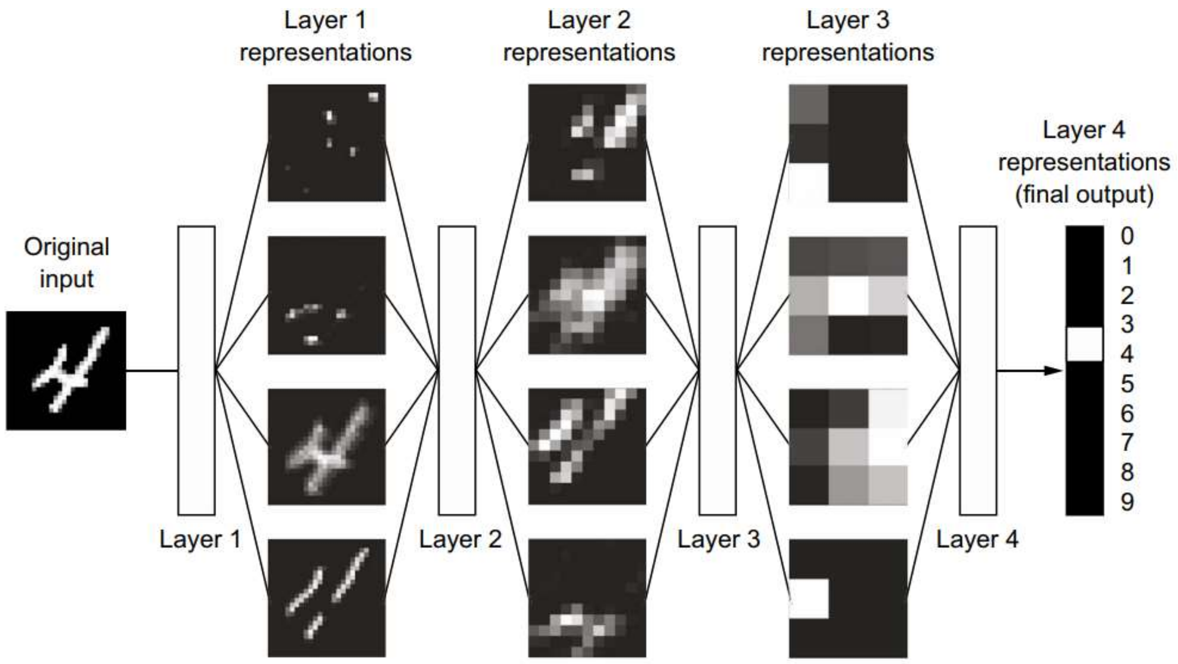
A Deep neural network for digit classification

Deep Learning Network



Data representation learned by a digit-classification model

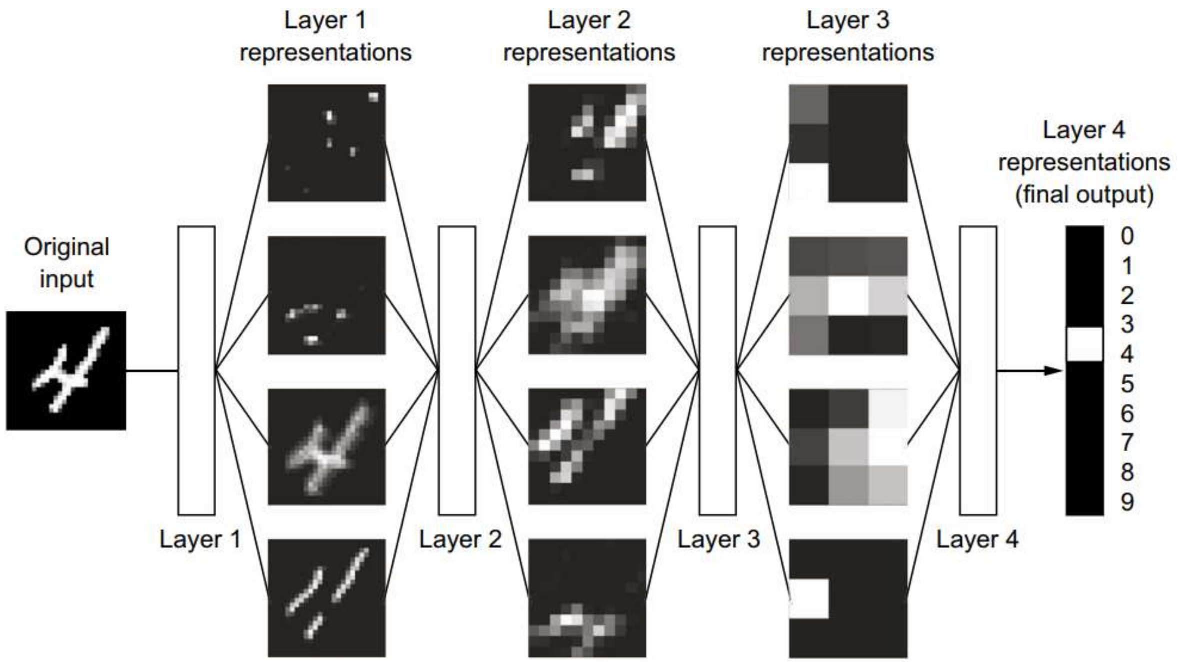
Deep Learning Network



Data representation learned by a digit-classification model

- The network transforms the digit image into representations that are increasingly different from the original image and increasingly informative about the final result.

Deep Learning Network



Data representation learned by a digit-classification model

- You can think of a deep network as a multistage *information distillation* process, where information goes through successive filters and comes out increasingly *purified* (that is, useful with regard to some task).

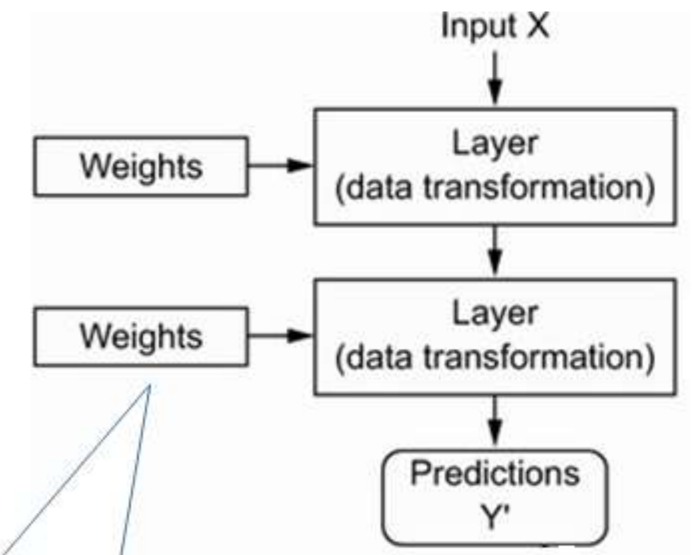
Deep Learning (DL)

- So that’s what deep learning is, technically:
  - a multistage way to learn data representations.
- It’s a simple idea—but, as it turns out, very simple mechanisms, sufficiently scaled, can end up looking like magic.

How Deep Learning works

- Now let’s look at how this learning happens, concretely.

How Deep Learning works

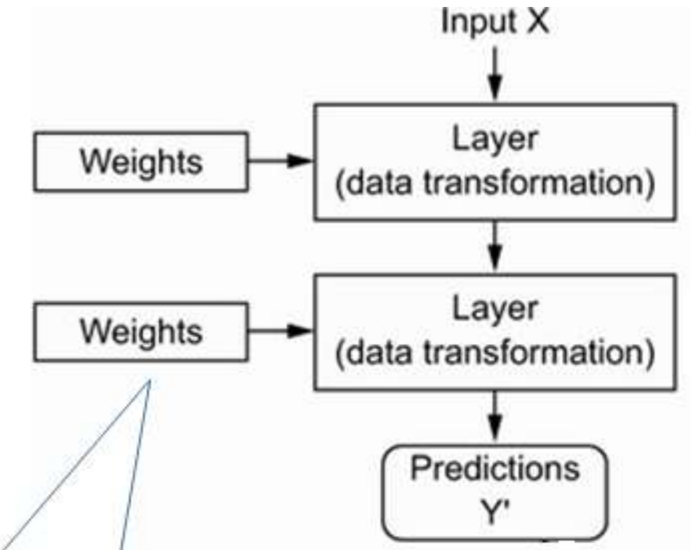


Goal: Finding the right values for the Weights

- The specification of what a layer does to its input data is stored in the layer's weights, which in essence are a bunch of numbers.
- In technical terms, the transformation implemented by a layer is parameterized by its weights.
- Weights are also sometimes called the parameters of a layer.

A neural network is parameterized by its weights.

How Deep Learning works



Goal: Finding the right values for the Weights

- In this context, learning means finding a set of values for the weights of all layers in a network, such that the network will correctly map example inputs to their associated targets.

A neural network is parameterized by its weights.

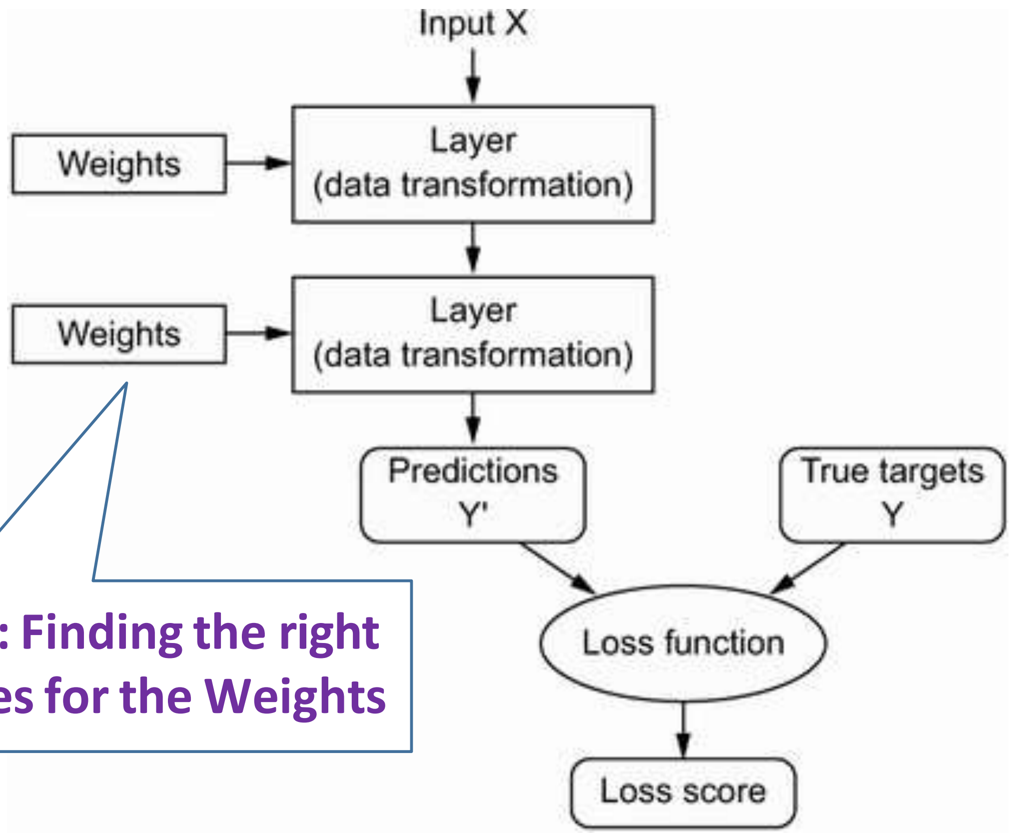
How Deep Learning works

- But…
- A deep neural network can contain tens of millions of parameters.
- Finding the correct values for all of them may seem like a daunting task, especially given that modifying the value of one parameter will affect the behavior of all the others!

How Deep Learning works

- To control something, first we need to be able to observe it.
- To control the output of a neural network, we need to be able to measure how far this output is from what we expected.

How Deep Learning works

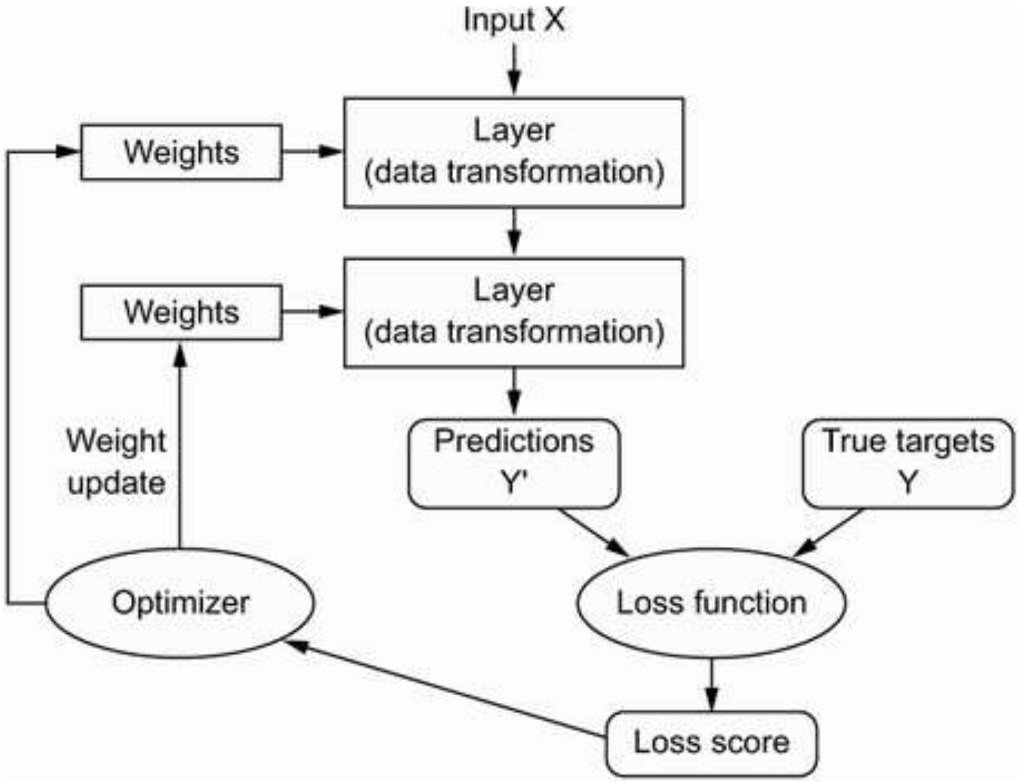


Goal: Finding the right values for the Weights

- This is the job of the loss function of the network, also sometimes called the objective function or cost function.
- The loss function takes the predictions of the network and the true target and computes a distance score, capturing how well the network has done on this specific example.

A loss function measures the quality of the network's output.

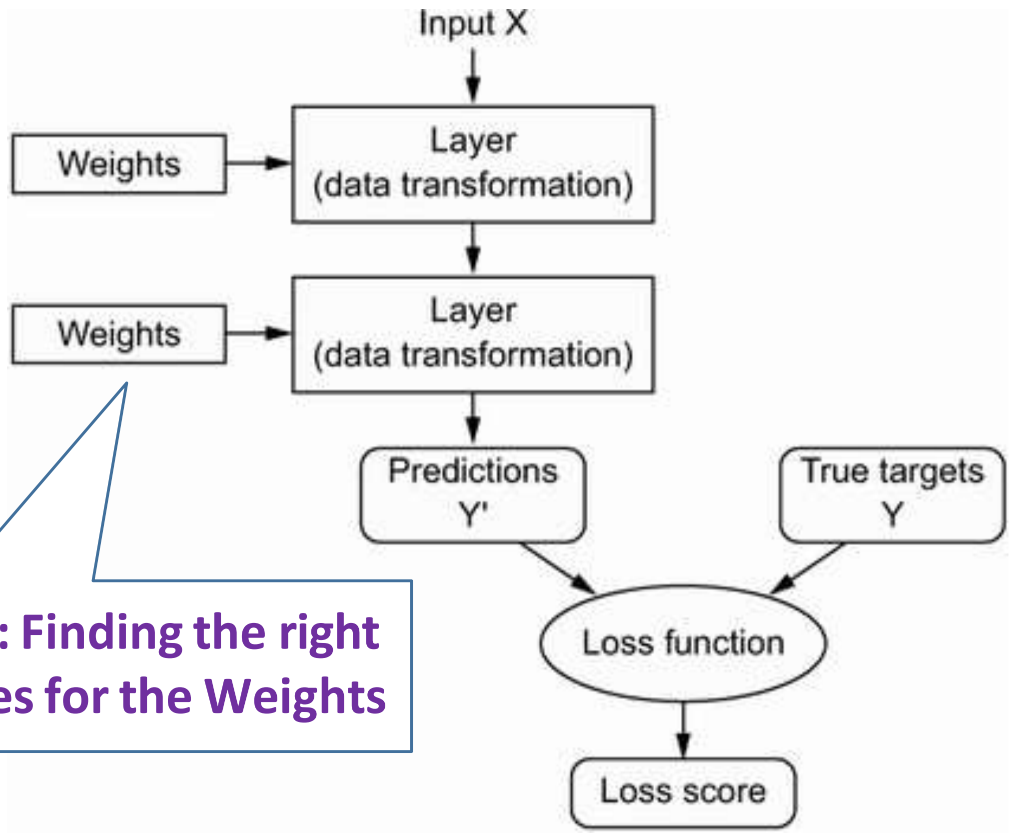
How Deep Learning works



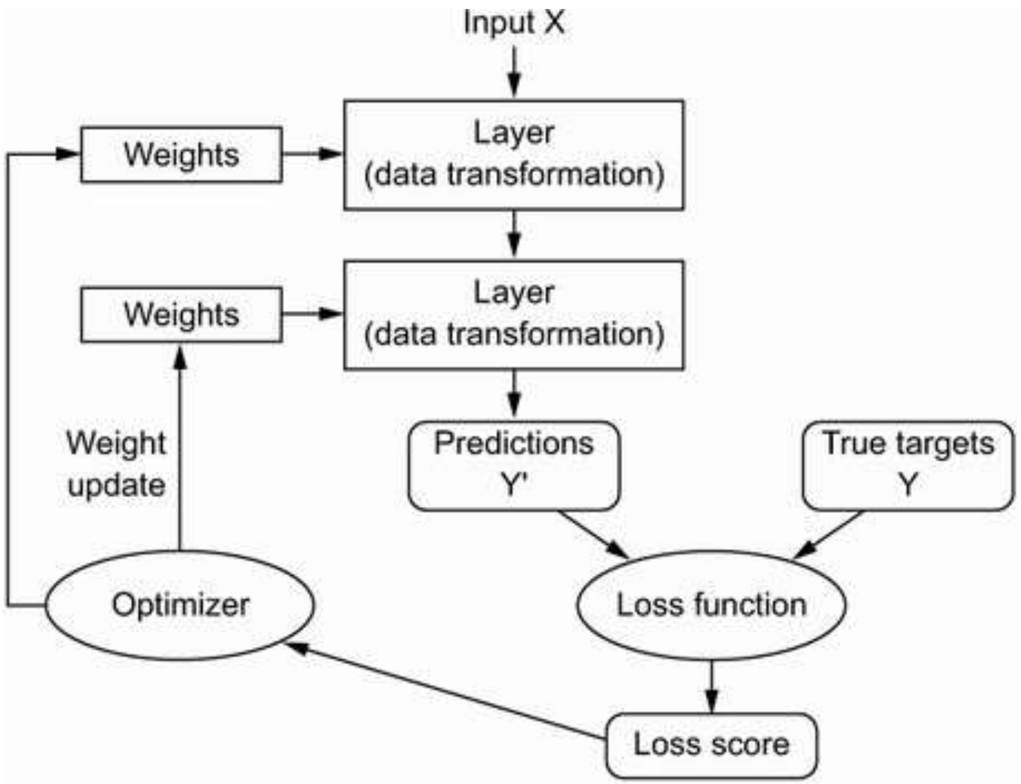
- The fundamental trick in deep learning is to use this score as a feedback signal to adjust the value of the weights a little, in a direction that will lower the loss score for the current example.
- This adjustment is the job of the optimizer, which implements what's called the Backpropagation algorithm: the central algorithm in deep learning.

The loss score is used as a feedback signal to adjust the weights.

How Deep Learning works



Goal: Finding the right values for the Weights



A neural network is parameterized by its weights.

How Deep Learning works

- Initially, the weights of the network are assigned random values, so the network merely implements a series of random transformations.
- Naturally, its output is far from what it should ideally be, and the loss score is accordingly very high.
- But with every example the network processes, the weights are adjusted a little in the correct direction, and the loss score decreases.

How Deep Learning works

- This is the training loop, which, repeated a sufficient number of times (typically tens of iterations over thousands of examples), yields weight values that minimize the loss function.
- A network with a minimal loss is one for which the outputs are as close as they can be to the targets: a trained network.
- Once again, it's a simple mechanism that, once scaled, ends up looking like magic.

What deep learning has achieved so far

- Near-human-level image classification
- Near-human-level speech transcription
- Near-human-level handwriting transcription
- 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

What deep learning has achieved so far

- Started applying deep learning with great success to a wide variety of problems that were thought to be impossible to solve just a few years ago:
    - automatically transcribing the tens of thousands of ancient manuscripts held in the Vatican's Apostolic Archive.
    - detecting and classifying plant diseases in fields using a simple smartphone.
    - assisting oncologists or radiologists with interpreting medical imaging data.
    - predicting natural disasters such as floods, hurricanes, or even earthquakes, and so on.

What deep learning has achieved so far

- Deep learning assists us in every activity and every field of human endeavor:
  - science
  - medicine
  - manufacturing
  - energy
  - transportation
  - software development
  - agriculture
  - even artistic creation

What makes deep learning different

- Deep learning completely automates what used to be the most crucial step in a machine learning workflow: feature engineering.
- Previous machine learning techniques—shallow learning—only involved transforming the input data into one or two successive representation spaces, usually via simple transformations such as high-dimensional non-linear projections (SVMs) or decision trees.
- But the refined representations required by complex problems generally can’t be attained by such techniques.
- As such, humans to manually engineer good layers of representations for their data. This is called feature engineering.

What makes deep learning different

- Two essential characteristics of how deep learning learns from data:
  - the incremental, layer-by-layer way in which increasingly complex representations are developed,
  - the fact that these intermediate incremental representations are learned jointly, each layer being updated to follow both the representational needs of the layer above and the needs of the layer below.
- Together, these two properties have made deep learning vastly more successful than previous approaches to machine learning.

Why deep learning? Why now?

- The two key ideas of deep learning for computer vision—convolutional neural networks and backpropagation—were already well understood by 1990.
- The Long Short-Term Memory (LSTM) algorithm, which is fundamental to deep learning for timeseries, was developed in 1997 and has barely changed since.
- So why did deep learning only take off after 2012?
- What changed in these two decades?

Why deep learning? Why now?

- Three technical forces are driving advances in machine learning:
  - Hardware
  - Datasets and benchmarks
  - Algorithmic advances

Hardware

- The NVIDIA Titan RTX, a GPU that cost $2,500 at the end of 2019, can deliver a peak of 16 teraFLOPS in single precision (16 trillion float32 operations per second).
  - That's about 500 times more computing power than the world's fastest supercomputer from 1990, the Intel Touchstone Delta.
- In 2007, NVIDIA launched CUDA, a programming interface for its line of GPUs.
- Google revealed its Tensor Processing Unit (TPU) project
- In 2020, the third iteration of the TPU card represents 420 teraFLOPS of computing power.
  - That's 10,000 times more than the Intel Touchstone Delta from 1990.

Data

AI is sometimes heralded as the new industrial revolution. If deep learning is the steam engine of this revolution, then data is its coal:

- User-generated image tags on Flickr, for instance, have been a treasure trove of data for computer vision.
- YouTube for video dataset.
- Wikipedia for natural language processing dataset.
- ImageNet dataset, consisting of 1.4 million images that have been hand annotated with 1,000 image categories is a key dataset for natural language processing.

Benchmarks

- Having common benchmarks that researchers compete to beat has greatly helped the rise of deep learning.

Algorithms

- Until the late 2000s, we were missing a reliable way to train very deep neural networks.
- As a result, neural networks were still fairly shallow, using only one or two layers of representations.
- The key issue was that of gradient propagation through deep stacks of layers.
    - The feedback signal used to train neural networks would fade away as the number of layers increased.

Algorithms

This changed around 2009–2010 with the advent of several simple but important algorithmic improvements that allowed for better gradient propagation:

- Better *activation functions* for neural layers
- Better *weight-initialization schemes*, starting with layer-wise pretraining, which was then quickly abandoned
- Better *optimization schemes*, such as RMSProp and Adam

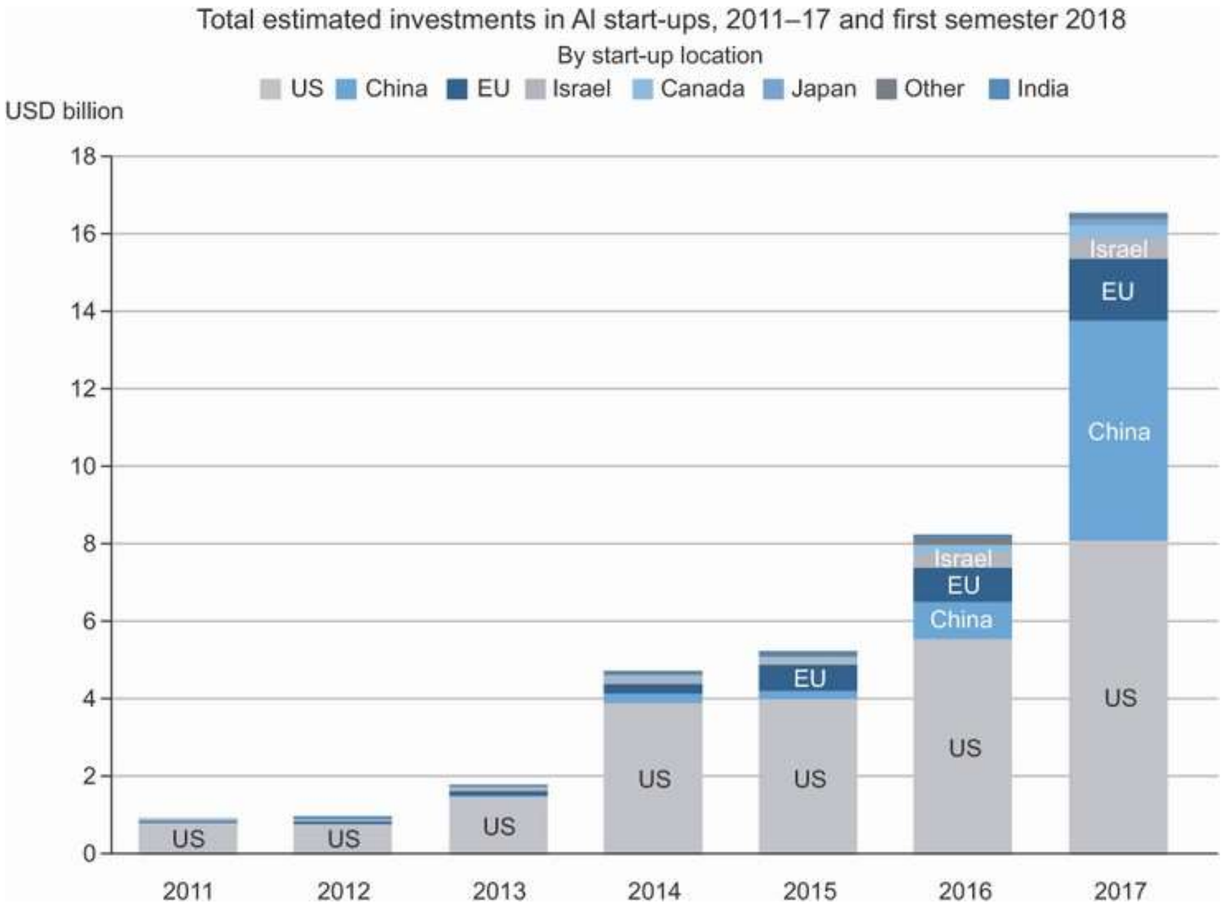
Only when these improvements began to allow for training models with 10 or more layers did deep learning start to shine.

Algorithms

Finally, in 2014, 2015, and 2016, even more advanced ways to improve gradient propagation were discovered, such as

- batch normalization,
- residual connections, and
- depthwise separable convolutions.

A new wave of investment in AI Startups



- In 2011, right before deep learning took the spotlight, the total venture capital investment in AI worldwide was less than a billion dollars.
- in 2015, it had risen to over $5 billion
- in 2017, to a staggering $16 billion

OECD estimate of total investments in AI startups (Source: http://mng.bz/zGN6)

AI Revolution

- Deep learning has several properties that justify its status as an AI revolution.
- These important properties can be broadly sorted into three categories:
  - *Simplicity*
  - *Scalability*
  - *Versatility and reusability*

AI Revolution

- *Simplicity*
  - Deep learning removes the need for feature engineering, replacing complex, brittle, engineering-heavy pipelines with simple, end-to-end trainable models that are typically built using only five or six different tensor operations.

AI Revolution

- *Scalability*
  - Deep learning is highly amenable to parallelization on GPUs or TPUs, so it can take full advantage of Moore’s law.
  - In addition, deep learning models are trained by iterating over small batches of data, allowing them to be trained on datasets of arbitrary size.

AI Revolution

- *Versatility and reusability*
    - Unlike many prior machine learning approaches, deep learning models can be trained on additional data without restarting from scratch, making them viable for continuous online learning—an important property for very large production models.
    - Furthermore, trained deep learning models are repurposable and thus reusable: for instance, it’s possible to take a deep learning model trained for image classification and drop it into a videoprocessing pipeline. This allows us to reinvest previous work into increasingly complex and powerful models. This also makes deep learning applicable to fairly small datasets.

AI Revolution

Deep learning is still a revolution in the making, and it will take many years to realize its full potential.