Introduction

In-Koo Cho

Emory University

December 1, 2020

Artificial intelligence

Francois Chollet [2018]: Deep Learning with Python, Manning Publications

> *Artificial intelligence is the effort to automate intellectual tasks normally performed by humans. (page 4)*

We have heard a lot about AI, machine learning and deep learning.

Diagram

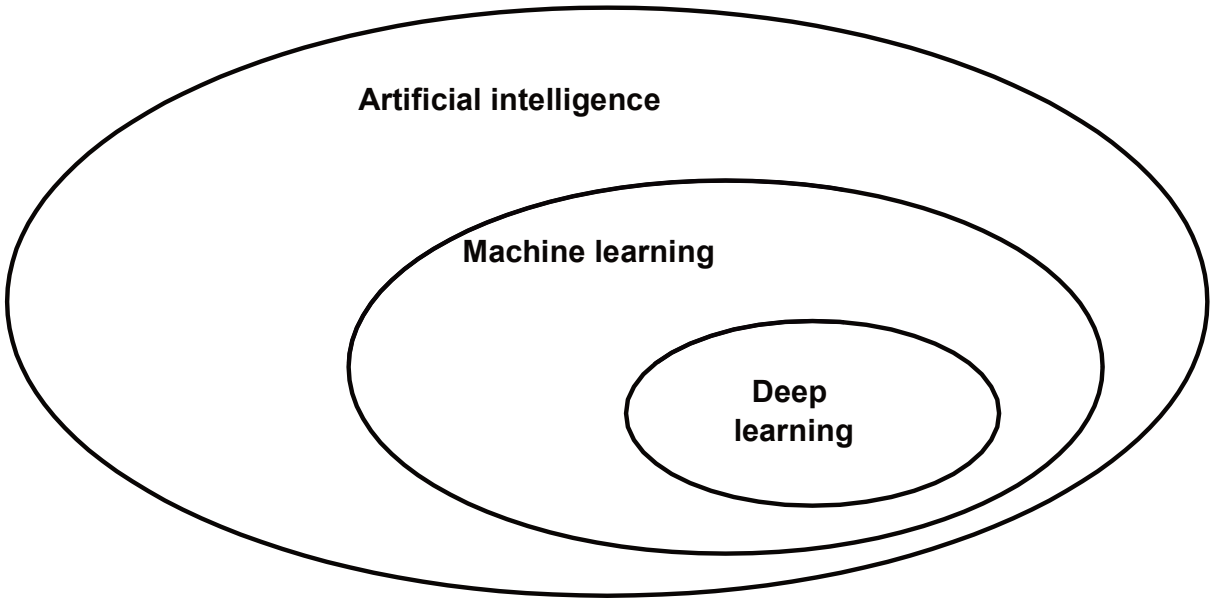

Figure 1.1 Artificial intelligence, machine learning, and deep learning

Beginning of Machine Learning

1. Lady Ada Lovelace invented the Analytical Engine, which is the first known general purpose mechanical computer. She said in 1843

   "The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. Its province is to assist us in making available what we are already acquainted with."

2. Alan Turing [1950]: "Computing Machinery and Intelligence," Mind 59, no. 236 (1950): 433-460, where he considered the remark as "Lady Lovelace's objection"
3. Turing explored whether general-purpose computers could be capable of learning and originality, and he came to the conclusion that they could.

Fundamental question

"Could a computer go beyond what we know how to order it to perform and learn on its own how to perform a specified task?" (Chollet, page 5)

While this question sounds like a conventional question in a computer programming class, the machine learning takes a fundamentally different view from the classical programming paradigm.

Classial programming

When we write a code, we specify the rule to process the data and produces the answer. For example, if you write a least square estimator, the program takes the data and processes the data according the rule to produces the forecast.

Machine learning

Instead of producing an answer (or forecast), the machine learning algorithm takes the data to train the rule by comparing the forecast from the answers. After training, the algorithm produces a rule, which will generates an answer on our behalf.

In the machine learing algorithm, we produce a program, instead of a forecast. Instead of answering a question (or making a forecast), the machine learning algorithm train the program using the data and comparing the outcome to the answers.

New paradigm

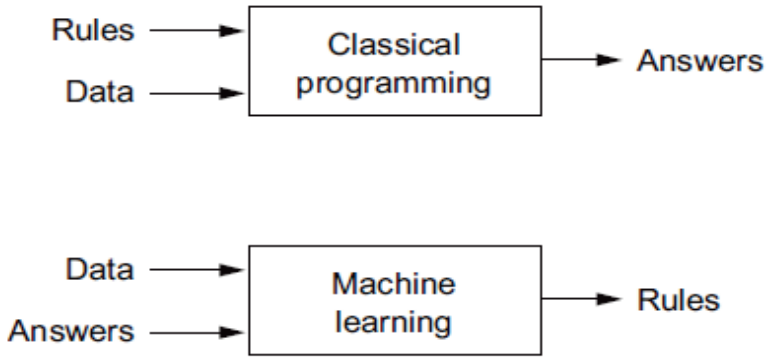

Figure 1.2 Machine learning: a new programming paradigm

Why recent hype?

To understand the recent hype of machine learning algorithm, we need to look at the history of AI.

Beginning

- Ada Lovelace [1843]: Intelligence Machine
- Alan Turing [1950]: “Can machine think?”

Post World War II

- Invention of a digital computer.
- Invention of (artificial) neural network by Marvin Minsky [1954] in his doctoral dissertation titled "Theory of neural-analog reinforcement systems and its application to the brain-model problem."
- Possibility of achieving human level general intelligence

Optimism

- Marvin Minsky claimed in 1967,

  Within a generation [. . .] the problem of creating 'artificial intelligence' will substantially be solved.

  In 1970, he predicted

  In from three to eight years we will have a machine with the general intelligence of an average human being.

- In the 1960s and early 1970s, several experts believed it to be right around the corner (as do many people today).

First Winter of AI

A few years later, the professional could not meet the expectations.

Researchers and government funds turned away from the field, marking the start of the *first AI winter*.

Expert System

Interest in AI revived, as *expert systems* show promise. The interest in the artificial intelligence spiked in early 80's, as the expert system shows a great promise to carry out a complex instruction on behalf of a human decision maker.

The expert system requires a complex code designed for a specific task. As the problem becomes more complex, the code becomes more complex. Different problems required different (and complex) codes.

Limits of expert systems

- **Complexity.**

Limits of expert systems

- **Complexity.** We need increasingly more complex code to carry out a complex task.

Limits of expert systems

- **Complexity.** We need increasingly more complex code to carry out a complex task.
- **Narrow scope.**

Limits of expert systems

- **Complexity.** We need increasingly more complex code to carry out a complex task.
- **Narrow scope.** A single code is designed to carry out a narrowly defined single task.

Limits of expert systems

- **Complexity.** We need increasingly more complex code to carry out a complex task.
- **Narrow scope.** A single code is designed to carry out a narrowly defined single task.

Both factors contributed to increasing cost of the expert system. As the interest waned, the funding cut followed. The interest in AI has died out. The 2nd Winter of Artificial Intelligence set in by early 1990's.

Surge of interest

In 2012, the success of Hinton's group in the yearly large-scale image-classification challenge ImageNet revived the interest in neural network and eventually led to the interest in deep learning (Chollet [2018])

Although NNs trained by *backpropagation* had been around for decades, and GPU implementations of NNs for years, fast implementations of NNs on GPUs were needed to progress on computer vision. In 2011, this approach achieved for the first time superhuman performance in a visual pattern recognition contest.

`https://en.wikipedia.org/wiki/Deep_learning`

Key points

- **Back progagation.**

Key points

- **Back progagation.** The program is trained by comparing the output from the code to the actual observation ("answer"). The feedback is an integrated part of any learning algorithm.

Key points

- **Back progagation.** The program is trained by comparing the output from the code to the actual observation ("answer"). The feedback is an integrated part of any learning algorithm.
- **Parallel.**

Key points

- **Back progagation.** The program is trained by comparing the output from the code to the actual observation ("answer"). The feedback is an integrated part of any learning algorithm.
- **Parallel.** Graphics Processing Unit process graphical information by many processors, which of which can handle only simple instruction. The processors operate in a parallel fashion to process the graphic information efficiently. Instead of building a complex instruction, we can achieve considerable computational capability by combining many simple processors.

Why exciting?

A modern machine learning algorithm addresses the two critical weakness of the expert system: high cost of writing complex codes for different tasks.

Why exciting?

A modern machine learning algorithm addresses the two critical weakness of the expert system: high cost of writing complex codes for different tasks.

- **Training.**

Why exciting?

A modern machine learning algorithm addresses the two critical weakness of the expert system: high cost of writing complex codes for different tasks.

- **Training.** A single (machine learning) algorithm can be trained to take care of a broad scope of task. Instead of writing different codes, the machine learning algorithm generates different predictor, following supervised training.

Why exciting?

A modern machine learning algorithm addresses the two critical weakness of the expert system: high cost of writing complex codes for different tasks.

- **Training.** A single (machine learning) algorithm can be trained to take care of a broad scope of task. Instead of writing different codes, the machine learning algorithm generates different predictor, following supervised training.
- **Boosting.**

Why exciting?

A modern machine learning algorithm addresses the two critical weakness of the expert system: high cost of writing complex codes for different tasks.

- **Training.** A single (machine learning) algorithm can be trained to take care of a broad scope of task. Instead of writing different codes, the machine learning algorithm generates different predictor, following supervised training.
- **Boosting.** A single (machine learning) algorithm build a code from the combination of simple codes, which is capable of carrying out complex task. By combining a weak predictor (slightly better than random guess), we can build a strong predictor (arbitrarily accurate predictor).

Adaptive boosting

Among many algorithms, Adaptive Boosting algorithm (AdaBoost) invented by Schapire and Freund [2012] is very much successful in applications, and generates substantial theoretical interest.

"Learning" in machine learning differs from "learning" in neural network learning. The result from neural network learning is an updated forecast (decision), but the result from machine learning is a forecating rule (classifier).

We often call a machine learning algorithm a statistical procedure to avoid confusion.

Equilibrium model

A rational agent instantaneously makes an optimal choice for any given decision problem.

Rationality is defined independently of the decision problem, and therefore, is an ex-ante concept. It is a behavioral assumption that applies uniformly to a decision problem.

Some problems are complex. Rationality assumes unlimited computational capability, which we find excessive. Investigation of various implications of limited computational capabilities follows.

Bounded rationality

- Human beings are not fully rational, because we have only limited computational capability.
- Abstract model of computers offers a useful tool to formalize the computational complexity, and to formulate how a decision maker copes with limited computational capability.
- An important implication of limited computational capability is the lack of perfect foresight. The forecast of a decision maker differs from the actual observations.
- A decision maker can learn from the errors, and updates the forecasts. Thus, the learning behavior arises naturally.

Computational cost

A human is designed to economize the energy to process information. Thus, we prefer a simple problem to a complex problem, if these two problems lead to the same payoff Rubinstein [1986] shows that the trade-off between the complexity cost and the payoff can lead to a dramatic difference from the prediction of an equilibrium model.

Rodolfo R. Llinas [2002]: "i of the vortex: From Neurons to Self," MIT Press

Imperfect foresight

It is too complex to model the actual law of motion accurately. An agent typically relies on a simpler, but manageable model, even if it is misspecified. Sargent [1993] surveyed studies on dynamic models in which an agent forms the belief as an econometrician, rather than a rational agent.

Bounded Rationality in Macroeconomics [1993] by Thomas J. Sargent, Oxford University Press

Statistical learning in economics

- New estimation method. Athey, Imbens et al.
- Classification of data to improve prediction. Fudenberg and Liang [2019]
- Statistical procedure. Al-Najjar [2009]; Al-Najjar and Pai [2014]
- Learning with Statistical Procedures and Misspecified Models. Cherry and Salant [2019]; Olea, Ortoleva, Pai and Prat [2019]; Spiegler [2018]; Eliaz and Spiegler [2016]

Outline

- Finite automaton and Neural network
  - Representation and basic concepts
    - complexity measure
    - machine game
  - Learning and analytic tools
- Machine learning
  - Basic concepts in statistical learning
  - Adaptive Boosting algorithm