Data science and AI: Are they the same thing?

(Answer: Not quite)

Aapo Hyvärinen

Gatsby Computational Neuroscience Unit
University College London
&
Department of Computer Science
University of Helsinki

HiData Kick-Off May 2018
What is Computer Science?

- “Study of the theory, experimentation, and engineering that form the basis for the design and use of computers” - Wikipedia

- But: “Computer science is no more about computers than astronomy is about telescopes”
  - Actually, about information and data (and knowledge?)

- Sometimes called *informatics*
  - but the word was trademarked in the US
What is Artificial Intelligence?

- “Study of ‘intelligent agents’: any device that perceives its environment and takes actions that maximize (...) its goals” - Wikipedia

- In practice today: Any really cool information processing

- Machine vision, search, games, translation, etc.
Brief History of AI

1940- Intelligence is logic (if-then, and, or)
1980- No: Logic is too limited. Intelligence requires imitating the brain.
1980- Intelligence requires learning
2000- Learning is just statistics. No need to imitate the brain.
2012- Intelligent processing is highly nonlinear, like the brain: Deep Learning
Artificial Intelligence is now based on machine learning

- Difficult to program intelligent operations
  - E.g. how do you recognize this is a cat?
- Perhaps easier to program a learning method
  - Then show it many cats (and other things)
Machine learning is (just?) statistics

- Elementary intelligence: linear regression
- Enables prediction
- Distinguish cats and dogs based on correlations of some features and class (cat/dog)?
  - Might work if we knew how to compute good nonlinear features (edges, corners, noses, legs, etc)
Deep learning of images

- Start of current AI boom in 2012
- Based on artificial neural networks
- Learns highly nonlinear features and processing from data
- Learns to recognize objects in photos
- Uses >1,000,000 images with labels

(Krizhevsky, Sutskever, Hinton, 2012)
What is Data Science, then?

- “Interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data” - Wikipedia

- Cf. “Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data.”

- Compared to statistics, data science adds emphasis on
  - Computation & algorithms
  - Data bases
  - Data transmission, networks
Three different kinds of machine learning

- **Supervised learning**
  - With input $x$ and output $y$, find relation
  - Regression generalized, pattern recognition

- **Unsupervised learning**
  - With just input $x$, try to understand its structure
  - Exploratory data analysis: clustering, factors/components

- **Reinforcement learning**
  - An *agent* acts in a world, getting input $x$, taking actions $a$, and obtaining reward $r$
  - Specific to AI
Autonomous agents are typical of AI

- For software agents, testbed is playing games
- They beat humans in video games, chess, Go, etc.
- Robotics may be the next big thing, possibly social robotics
Importance of human interpretation is typical of Data Science

Decompose brain signals into components:

Hyvärinen, Ramkumar, Parkkonen, Hari, 2010
Data Science vs AI

DATA SCIENCE
- Data bases
- Information Networks
- Interpretation

AI
- Computation & Algorithms
- Neural networks
- Big Data
- Agents/Rbots
- Cognitive Science

STATISTICS
- Mathematics
- Probability
- Real data
- Testing
Summary

• Are Data Science and AI the same thing?
  – Not quite!

• AI is ultimately about intelligent agents
  – But it needs a lot of data analysis to get there

• Data Science often aims at human insight
  – “From Data to Knowledge”? (former Academy CoE)

• Still, any borders are very fuzzy and plastic
  – Everything above can be published as machine learning