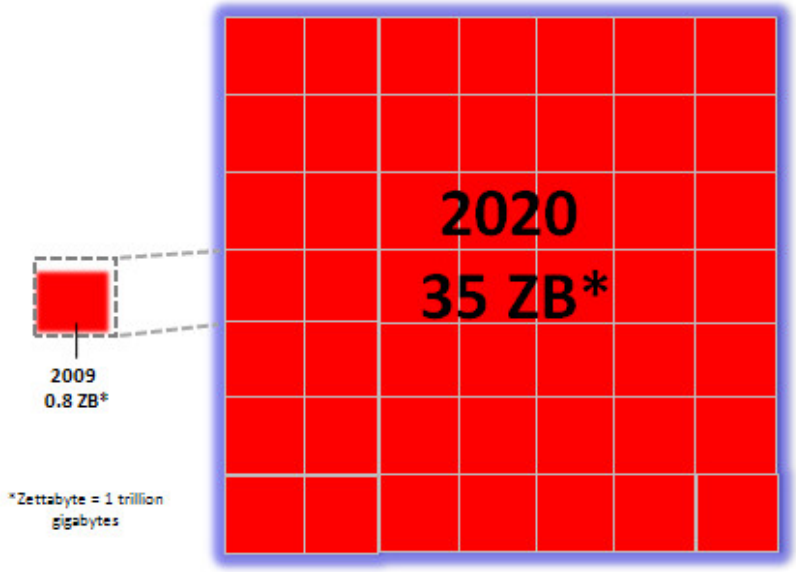
Machine learning

TM the Great Unread

DTL|Digital Arts Initiative
Interacting Minds Centre|Aarhus University

July 29, 2016

The emerging gap

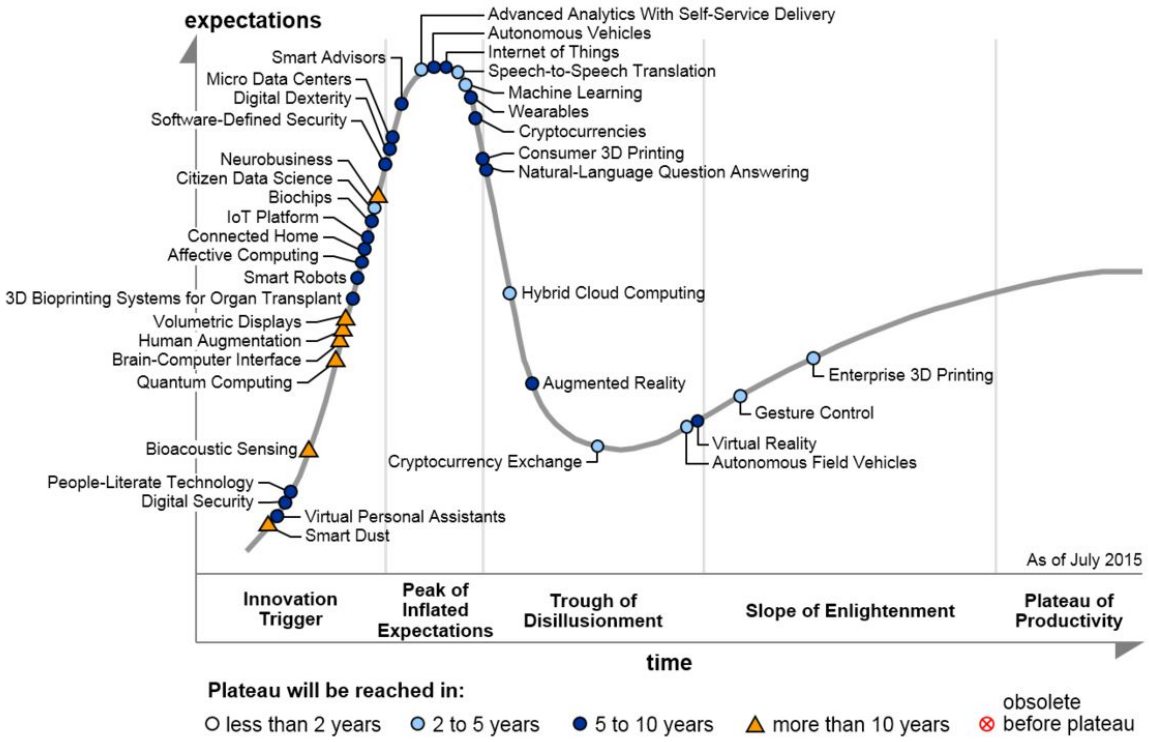


"We are drowning in information and starving for knowledge." (Roger)

Between 2010 and 2020 the amount of digital information will grow by a factor of 44, but the staffing and investment to manage it will grow by a factor of just 1.4 (IDC)

Task automation and intelligent machines are becoming exceedingly important

ML is the new black



Due to the explosion in data and measurements we can make of the world, ML is rapidly becoming (one of) the most important areas within computing science

Machine learning

We need statistical procedures for extracting important patterns and understanding what the data say $\rightarrow$ procedures that learn from data

Machine learning is about getting a computer to learn (extract patterns or structure from data) without explicit instructions. **Learning** means to solve a problem by generalizing from past experience (example data)

Need lots of data Fundamentally exploratory approach ("letting the data speak") that has become possible due to available data and computing power

Learning

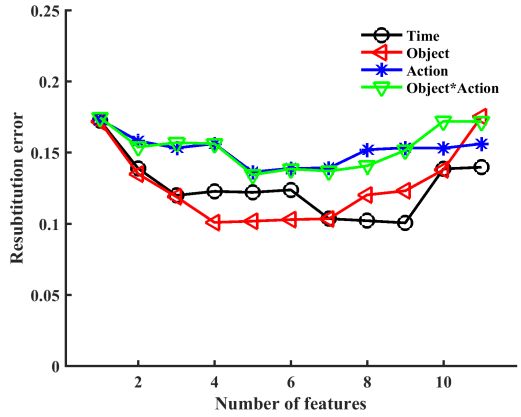
Slightly more technical way of understanding learning as
representation $\rightarrow$ *evaluation* $\rightarrow$ *optimization*

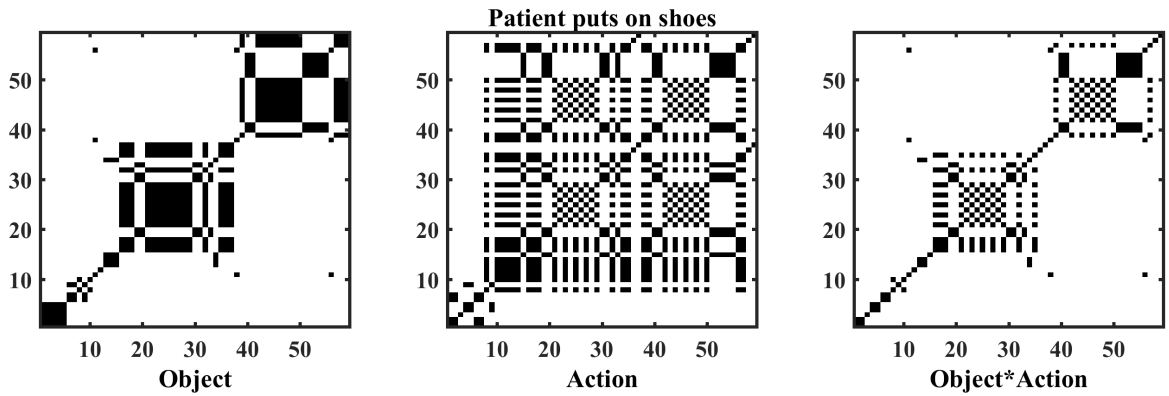
Two kinds of learning problems:

- **Supervised** goal is to predict the class (or numerical value of an outcome measure) based on a number of features[1]
- **Unsupervised** there is no class (outcome measure), and the goal is to describe[2] the associations and patterns among a set of features

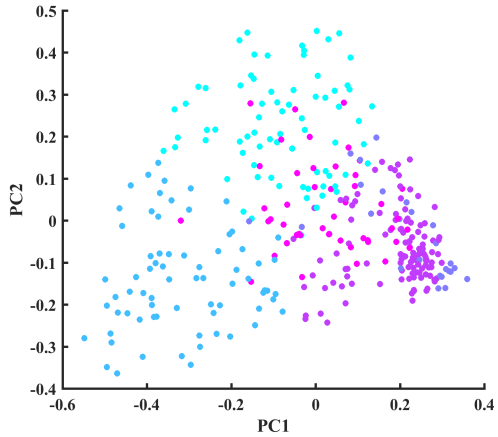
[1] data relevant to the task at hand (e.g., terms frequencies)
[2] data-driven exploration

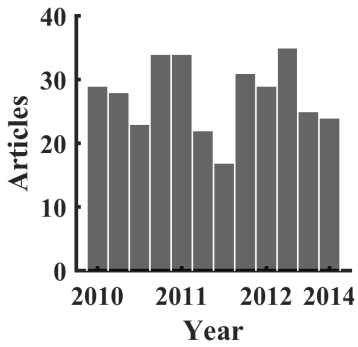
Supervised learning in the humanities

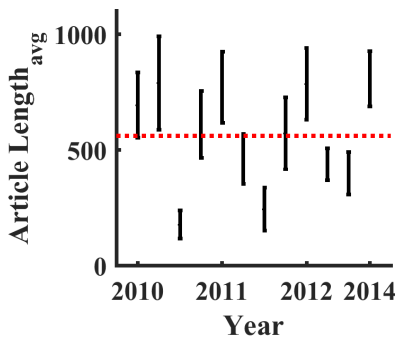


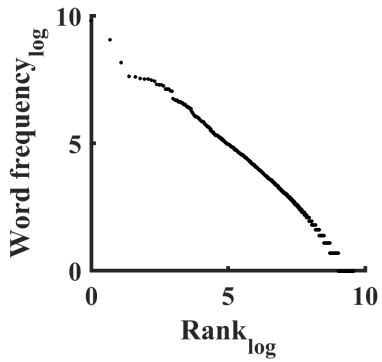


Unsupervised learning in the humanities





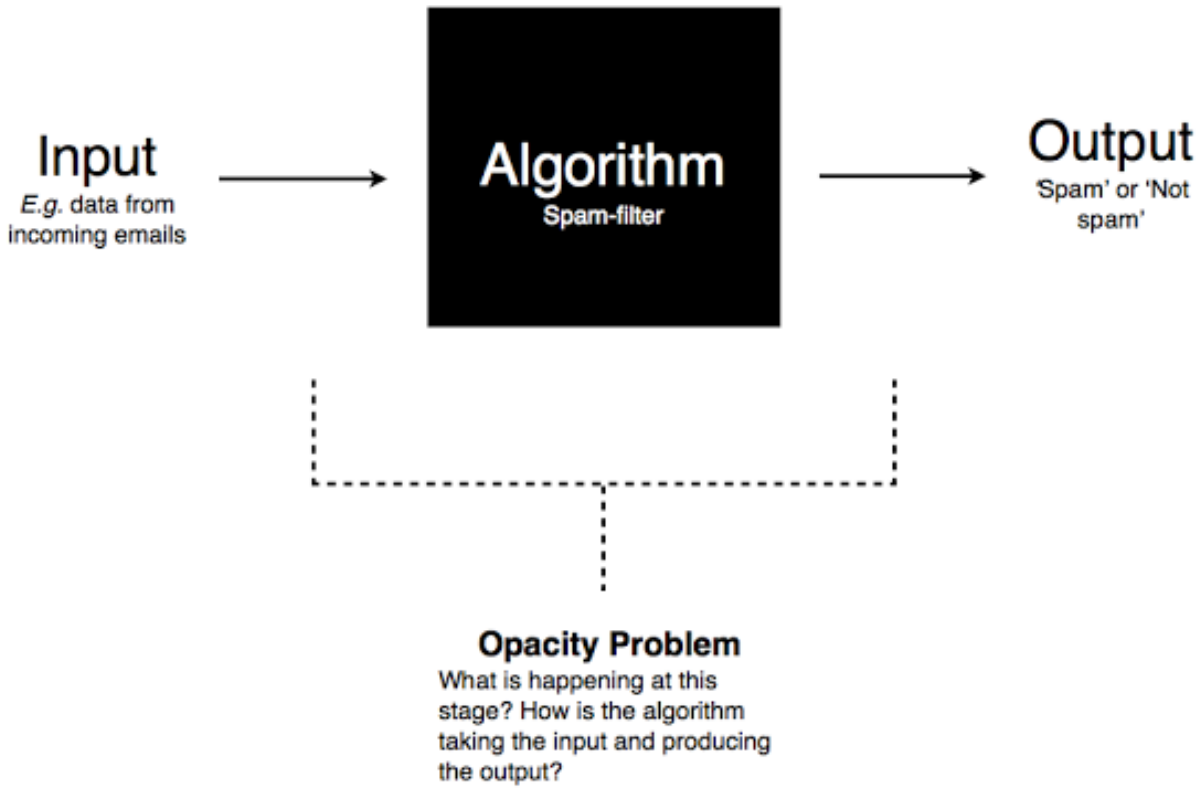




Issues

- The **black box**
- Applying predictive analytics to culture
- **Overfitting**
- Correlation does not imply causation

Black box methods



Overfitting