Critical Reasoning
Lecture-Seminar 8
Data Science (Part I)

Dr. Ioannis Votsis
NCH, Philosophy Faculty

ioannis.votsis@nchlondon.ac.uk
www.votsis.org

Introduction

What is data science?

- Data science is both an activity and a field of education.

  Activity: Data collection/interpretation/analysis via automated and manual statistical methods to posit/evaluate models.

  Education: The study of data collection/interpretation/analysis statistical/machine-learning methods and model selection.

- The models codify knowledge in a given area and are used to generate specific decisions under specific conditions.

- The application of data science to problems is so wide-ranging that it benefits both academic and business research.

Why do I keep hearing about data science?

- There is currently a huge demand for data science:

  "Demand for workers with specialist data skills like data scientists… has more than tripled over five years" (Royal Society Report published 09/05/19).

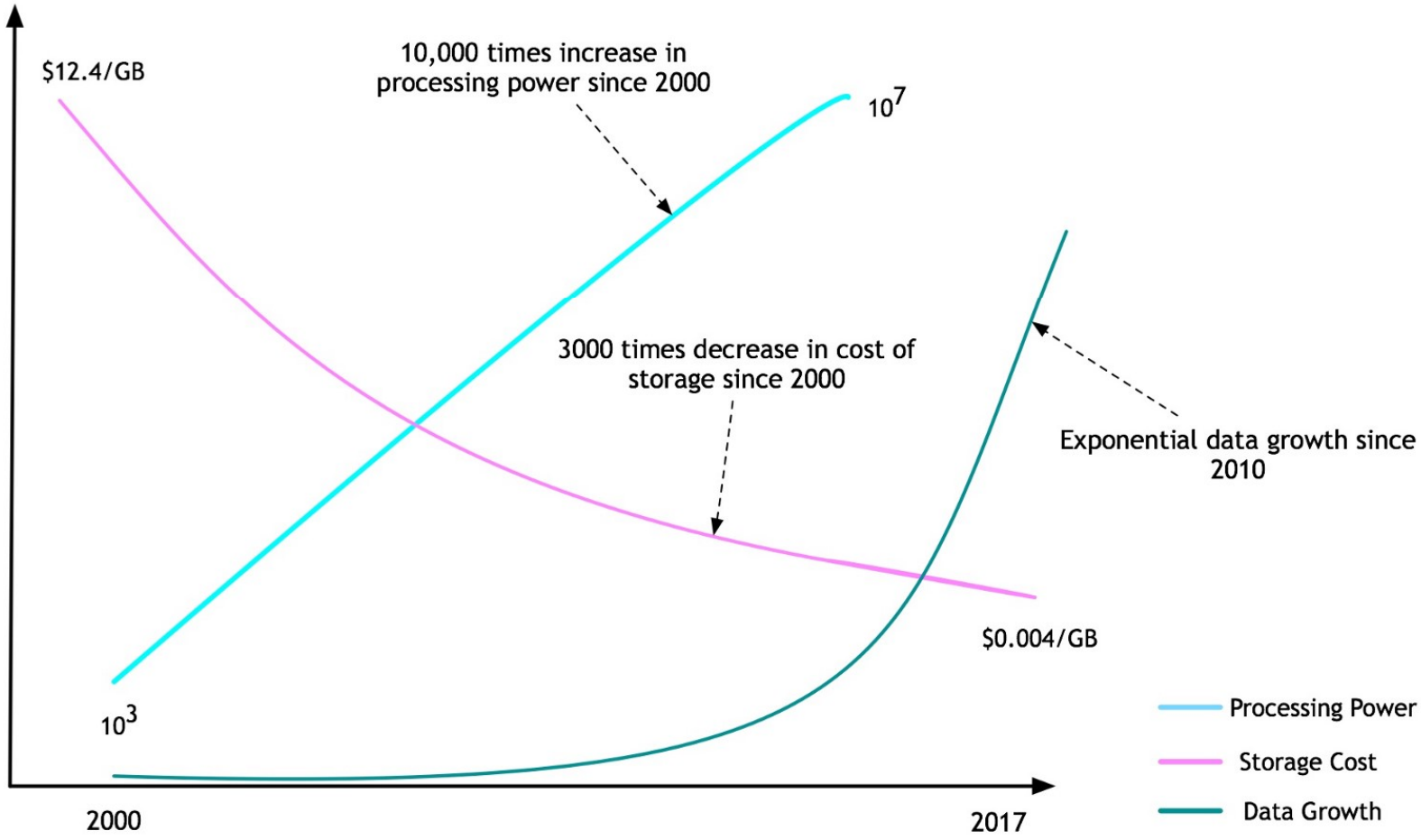
- That's no wonder as in recent years:

  * there are many more smart devices collecting data and, as a result, an incredible amount of data is available

  * computational power has increased considerably

Growth set to be stifled

- The trouble is that data growth seems to be outpacing computational power growth (see medium.com).



Why care about data science?

- Because practicing it has been increasingly beneficial for both public and private organisations.

  Examples: Google, Amazon, Netflix and governments use data to predict behaviour and even to manipulate it.

- Why should *you* care?

  * It's a powerful tool that can supplement your existing skills – this area is now called 'digital humanities'.

  * Employability (lots of well-paid and future-proof jobs).

Constituent skills

- Consider the following description, an online-only professional certificate offered by Harvard hosted via edX.

  "The HarvardX Data Science program prepares you with the necessary knowledge base and useful skills to tackle real-world data analysis challenges. The program covers concepts such as **probability**, **inference**, **regression**, and **machine learning** and helps you develop an essential skill set that includes R **programming**, **data wrangling** with dplyr, **data visualization** with ggplot2, **file organization** with Unix/Linux, **version control** with git and GitHub, and **reproducible document preparation** with Rstudio".

  https://www.edx.org/

UCS-NCH Collaboration

- University Centre Shrewsbury partnered with us to develop a conversion MSc in Data Science.

- This endeavour is funded by the Institute of Coding and seeks to increase the participation of underrepresented groups.

- Envisioned compulsory modules:

  * Fundamentals of Data Science
  * Data Management and Transformation
  * Principles of Machine Learning
  * Data and AI Ethics
  * Advanced Machine Learning
  * Business Development, Consultancy and Entrepreneurship
  * Work placement project and dissertation

Northeastern and NCH

- As a result of our integration into the Northeastern network, we have re-aligned our short and long-term goals.

- To be precise, the plan is to bring Data Science minors to all NCH degree programmes from Sept. 2020. For example:

  Philosophy with Data Science
  * Existing Philosophy Modules
  * New Philosophy Modules (Data and AI Ethics, Minds and Machines, Philosophy of Information and Computing).
  * Data Science Modules (Principles of Computing, Programming with Data, Foundations of Data Science)

  NB: At this point, the list is merely indicative.

NCH's own MSc Programmes

- We are also looking to introduce:

  * MSc in AI with a Human Face (September 2020)

  * MSc in Digital Humanities (September 2021)

  NB: We are hoping to get OFS funding for both of these.

- The two MScs would make use of the aforementioned modules as well as introduce additional new ones.

Data Science: An Example

Big data

- The kinds of data that can be collected and analysed can vary dramatically. For example:

  * mouse clicks
  * web searches
  * purchasing behaviour
  * contents of tweets
  * GPS logs
  * reaction times
  * heartbeat sequences
  * vibration activity

- To use machine learning, particularly deep learning, we need vast amounts of data.

Big data and predictions

- Google Flu Trends (GFT) is now a defunct project (2008-2015) that was once used to estimate flu outbreaks.

- The basic data employed was web searches of related terms in certain areas of certain countries.

  Sample words: flu, fever, cough, sore throat, stuffy nose, cough syrup, cold and flu capsules, flu jab and lozenges.

- The project enjoyed some successes, e.g. the prediction of flu outbreaks, 2 weeks ahead of CDC surveillance program.

- The project was abandoned primarily because GFT's predictive power was insufficient and inconsistent.

GFT

google.org Flu Trends

Google.org home

Dengue Trends

Flu Trends

Home

Select country/region

How does this work?

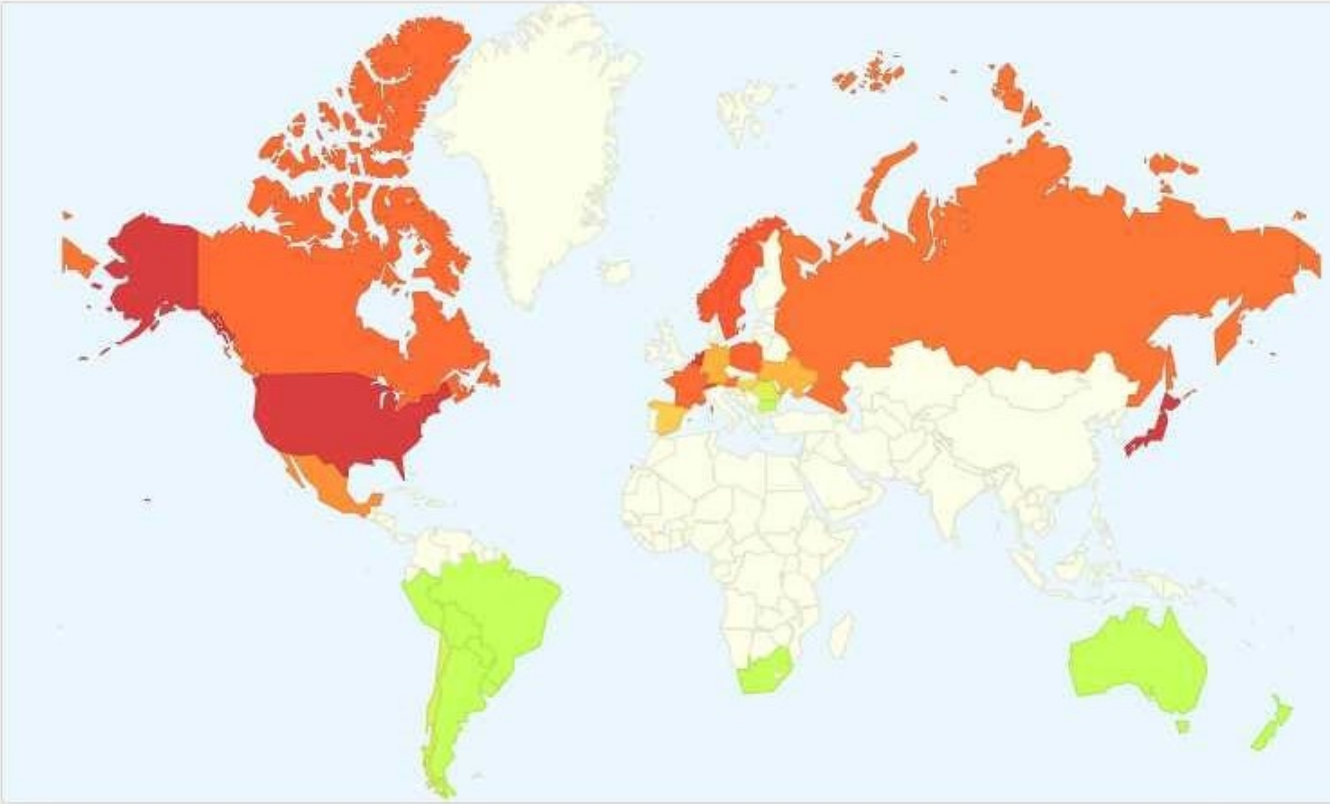
FAQ

Flu activity

- Intense
- High
- Moderate
- Low
- Minimal

Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. Learn more »



Download world flu activity data - Animated flu trends for Google Earth - Compare flu trends across regions in Public Data Explorer

Statistics automated (or something more)?

- Data science does not merely involve the use of human statisticians but also automates it.
- That's where machine-learning (ML) comes into the fore. ML algorithms are very good at detecting patterns.
- They use well-known statistical techniques like regression analysis to achieve these results.
- As a result, some people have accused machine-learning as being merely 'glorified statistics'.

Artificial Intelligence

What is artificial intelligence?

- Like data science, artificial intelligence (AI) is both an activity and a field of education.

- **Activity**: AI is the ability of an artificial unit to perform *intelligent* tasks in a(n) (semi-/)automated fashion.

- **Education**: The study of artificial systems in areas such as natural language processing, reasoning, sensors and robotics.

Levels of artificial intelligence

- A distinction is often drawn between two kinds or forms of AI: weak and strong.

  Weak AI: Artificial units that perform well in certain areas, e.g. playing games, solving equations or offering diagnoses.

  Strong AI: Artificial units that perform well in all areas and may even be said to be persons / have a consciousness.

- The holy grail in this field is to create units that reach or surpass general-level or human-level intelligence.

- The surpassing of human-level intelligence is known as superintelligence – see, Bostrom (2014), for a discussion.

GOFAI

- Initial attempts (50s-80s) to produce artificially intelligent systems focused on the rule-like manipulation of symbols.
- That is, they mimicked the way we prove theorems in mathematics and logic.
- John Haugeland has branded this approach 'good ol' fashioned AI' (GOFAI).
- The approach was not without success but it stalled in 70s and 80s, leading to what many called the 'AI winters'.
- Successes: expert systems, computer-generated players in games and some rudimentary robotics, etc.

Expert systems

- Expert systems were developed to mimic the decisions taken by human experts.

- Each such system had two key components: (i) an inference engine (logic and heuristics), (ii) a knowledge base.

- Some of the most successful expert systems were used to provide medical diagnoses.

Example:
If symptom = “weight loss” and “coughs blood” then they probable have lung cancer. (Knowledge Base)
Patient A has weight loss + coughs blood. (input datum)

Patient A probably has lung cancer (output – diagnosis).

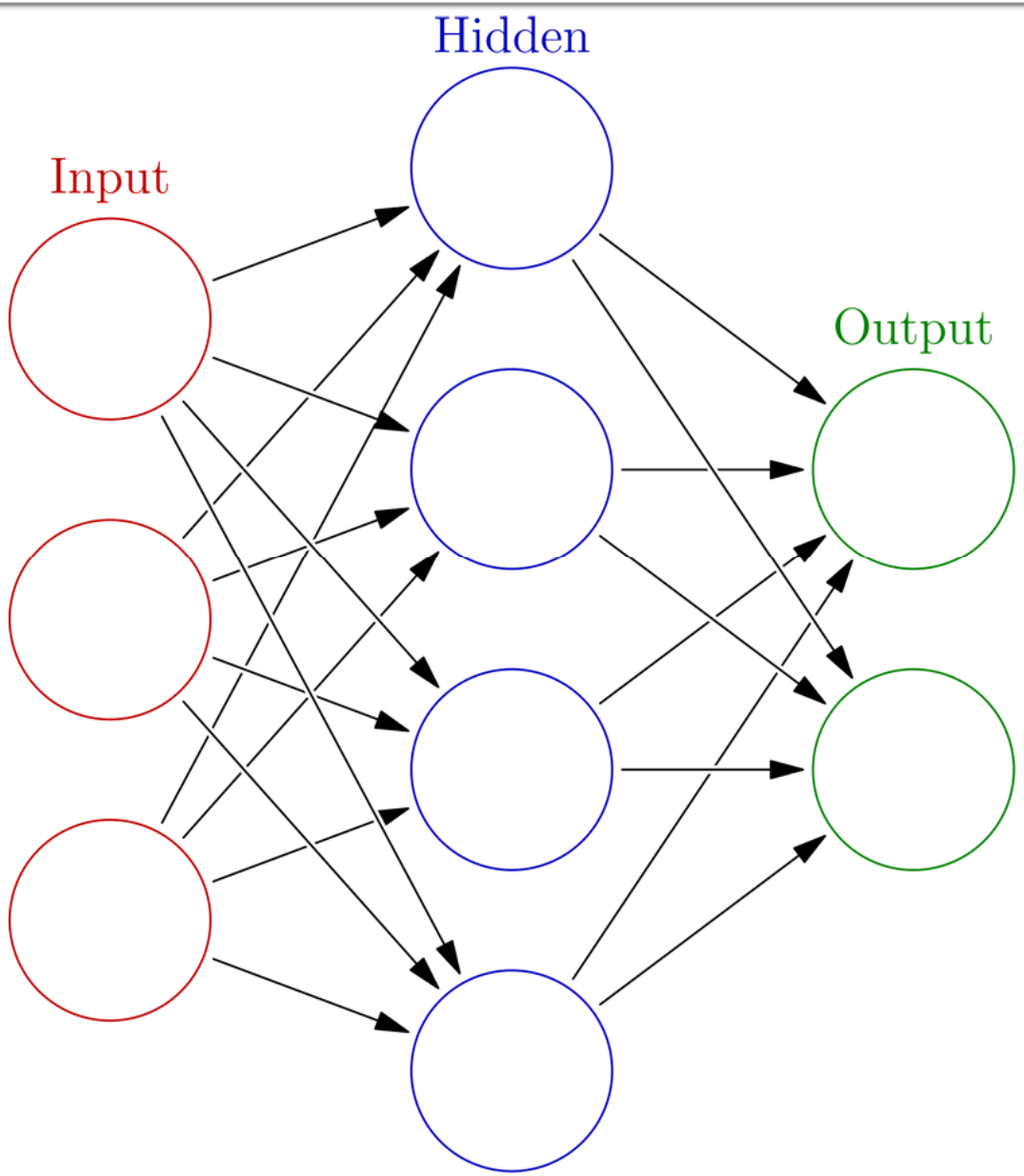
Expert systems: Example

- MYCIN was designed to diagnose blood and meningitis infections and to recommend apt drug treatments.
- In 1979, to test its efficacy, it was pitted against nine human doctors.
- Both humans and machine had to judge 10 real cases and recommend an appropriate drug treatment.
- Eight *other* human experts rated the recommendations without knowledge of which came from the expert system.
- MYCIN received the highest marks!

Neural Nets

- A radically different approach to artificially intelligent systems is via neural networks, or, *neural nets* for short.
- These mimic the interconnected web of neurons, also known as ‘neuronal networks’, that exists in the brain.
- The first-ever neural net was implemented by Belmont Farley and Wesley Clark at MIT in the 1950s.
- Neural nets went through a period of stagnation until interest picked up in the late 1980s.
- Today, neutral nets are the most widely used approach in artificial intelligence.

A simple feed-forward neural network

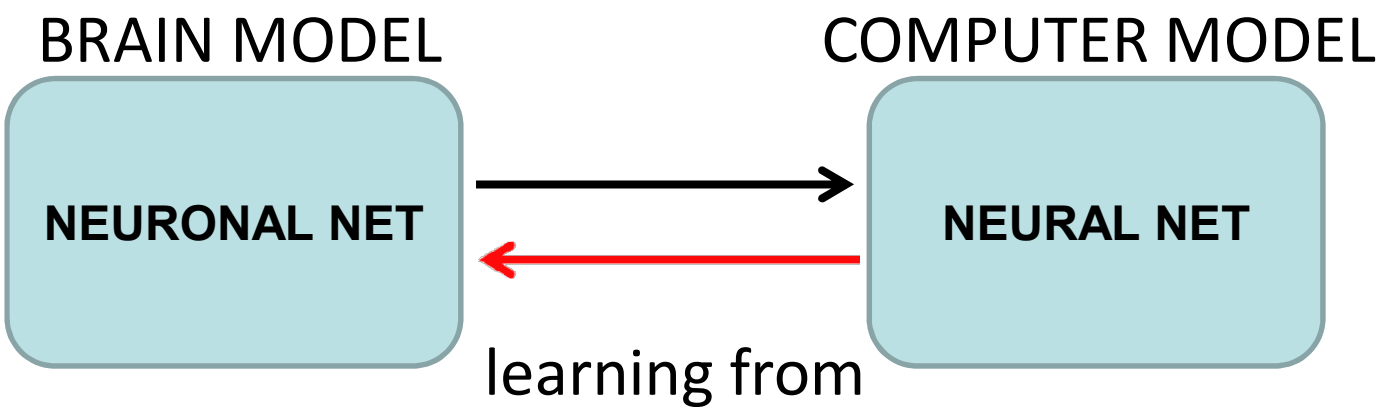


Connectionism

- In cognitive science, there are those who think that we can use computing as a model of the mind/brain.

- This approach is called 'the computational theory of mind' as the mind is taken to behave like a computer.

- Those who advocate the use of neural networks to understand the brain are known as '**connectionists**'.



Machine learning and deep neural networks

- ‘Machine learning’ and is a sub-branch of AI where an algorithm learns as much as possible on its own.

- The learning is not done through experience, (like human learning), but from data.

- Nowadays, most machine learning is conducted by neural nets. This includes deep learning models.

- They are thus called because they involve neural nets with multiple layers of hidden nodes.

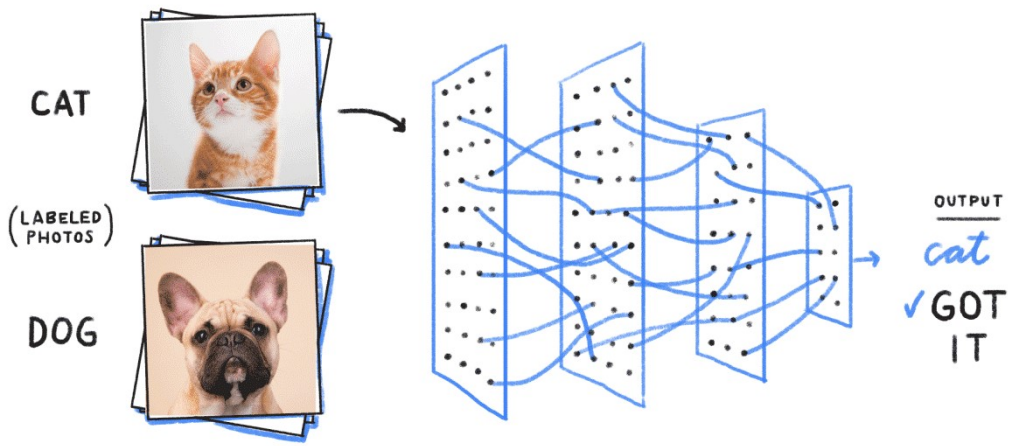
- Machine learning comes in two main forms: supervised and unsupervised.

Supervised learning

- **Supervised learning**: The algorithm is trained with labelled data and then its output is checked in the test phase.

- Labelling the data means that we know the desired result. The aim is to get the algorithm to produce the right labels.

- This form of learning is useful in *classification* and *curve-fitting* tasks.

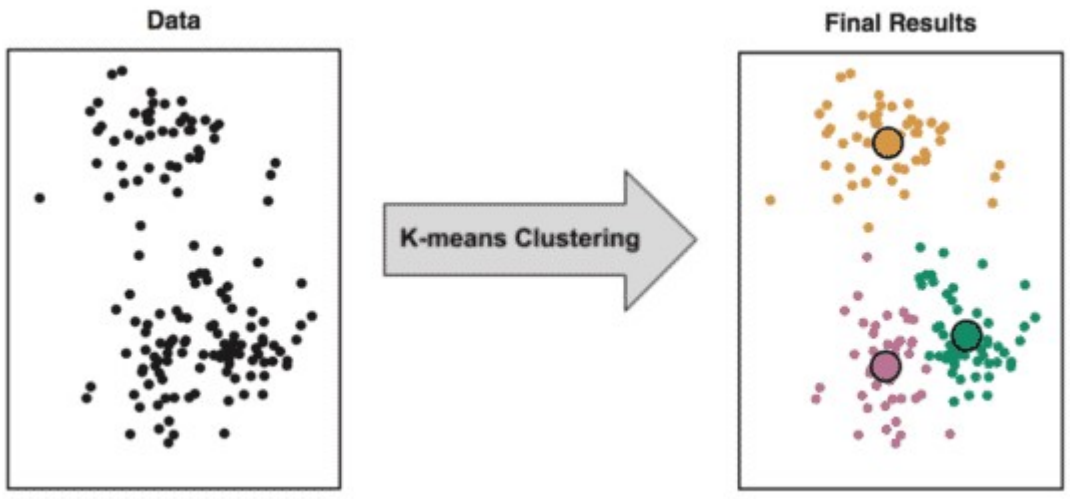


Unsupervised Learning

- **Unsupervised learning**: The algorithm is trained but the data employed is not labelled.

- That means we do not know the desired result and want the algorithm to produce its own pattern from the data.

- This form of learning is useful in *clustering* tasks.



A brief list of accomplishments

- There are a number of things already accomplished with AI. This includes beating world champions in games:

  * Chess: Deep Blue Vs. Kasparov (1996)
  * Go: AlphaGo Vs. Lee Sedol (2016)
  * Dota 2: OpenAI Five Vs. Team OG (2019)

- Beyond games, AI has been able to do well in a number of professional tasks, including science.

  * Autopilots in aeroplanes, drones and cars
  * Case Cruncher vs. 100 London-based lawyers (2017)
  * MIT machine discovered halicin as an antibiotic (2019).

The End