Introduction

Lê Hồng Phương
Data Science Laboratory, VNU Hanoi
<*phuonglh@hus.edu.vn*>

September 9, 2020

General Information

- Machine Learning, a practical approach
- 3 credits: 2*15 lecture sessions + 2*15 lab sessions
- Prerequisites:
  - Mathematics: linear algebra, multivariate calculus, optimization, probability and statistics
  - Computer science: data structure and algorithm, programming with a high-level language (such as Python, Java/Scala, C/C++)

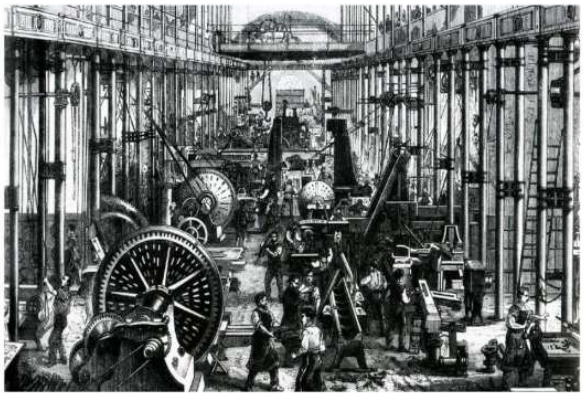
General Information

- Website of the course: *https://classroom.google.com/u/2/c/MTYwNDY1MjQyMTAy*
- Lecture notes, data sets, assignments and other information are available on the website. Check it out before each weekend!
- Programming language for the course: **Julia**. This language is simpler and faster to learn and use than **Python** or **Java**.

The Past Three Centuries

18th

the great mechanical systems and the industrial revolution



19th

the age of steam engine

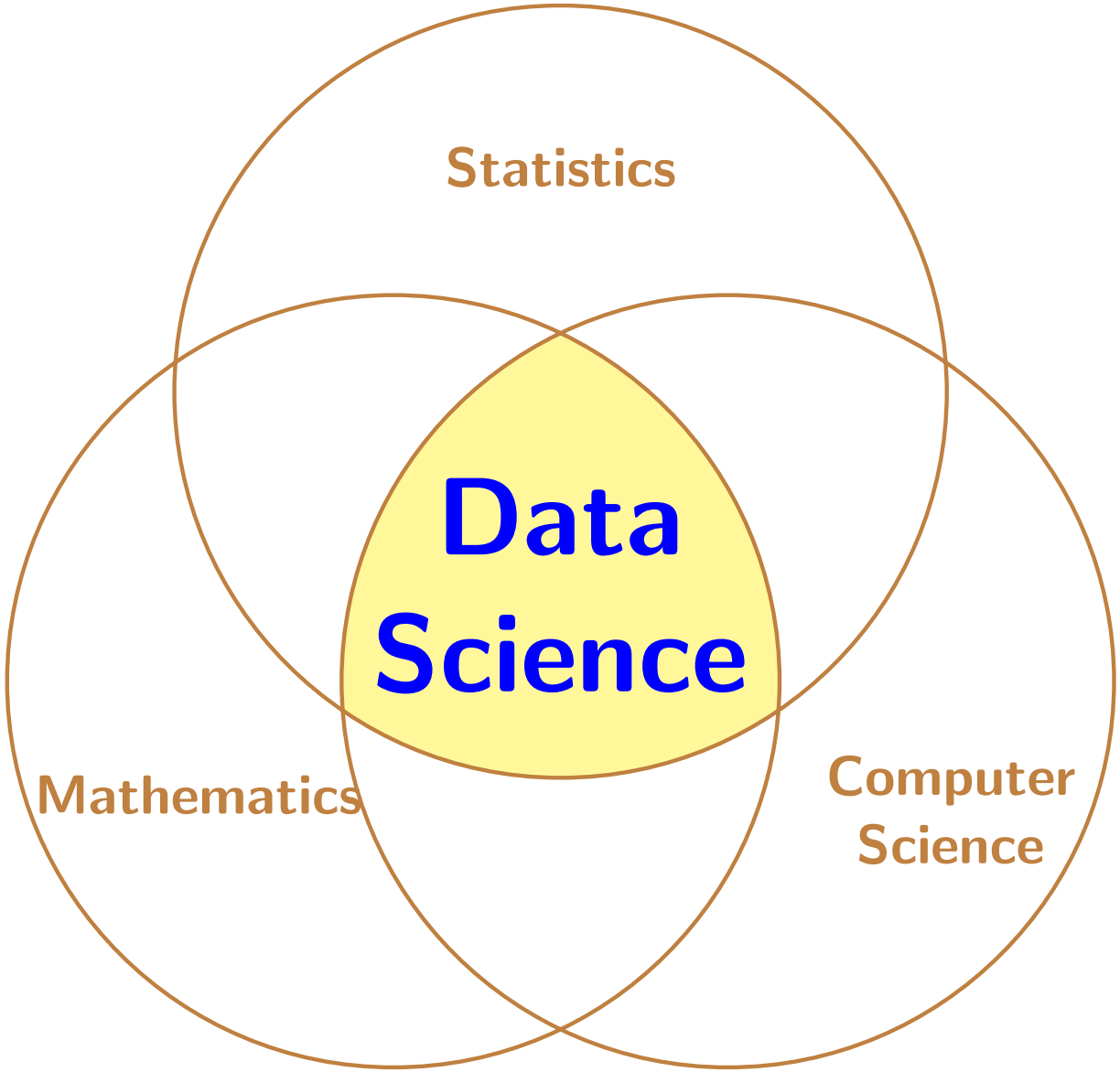
20th

electricity, radio, television, telephone, computer and Internet

Data Science

- An interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from data in various forms, both structured and unstructured.
- "The sexiest job of the 21st century" (Harvard Business Review, 2012)
- Data Science employs techniques and theories from many fields.

Data Science



Data Science

- Jim Gray (a Turing award winner) considered Data Science as the fourth paradigm of science (empirical, theoretical, computational, data-driven)
- IEEE Data Science and Advanced Analytics (DSAA):
- A core driver of the next-generation science, technologies and applications
- Is driving new researches, innovation, profession, economy and education across disciplines and across domains *http://www.dsaa.co*

Data Science Laboratory at HUS

General Information	• Officially established in December 26, 2017 • Head: Dr. Lê Hồng Phương • Website: http://mim.hus.vnu.edu.vn/dsl
Missions	• Industry relations and technology transfer • Fundamental and applied research in data science • Computer labs for instructional and training use
Research & Development Teams	1. Natural Language Processing 2. Image Processing 3. Mathematical Finance 4. Optimization and Operations Research 5. Applied Statistics and Probability 6. Cryptography and Data Security 7. Dynamical Systems and Biological Mathematics

Machine Learning

Machine Learning (ML) is at the heart of Data Science

Tom M. Michell

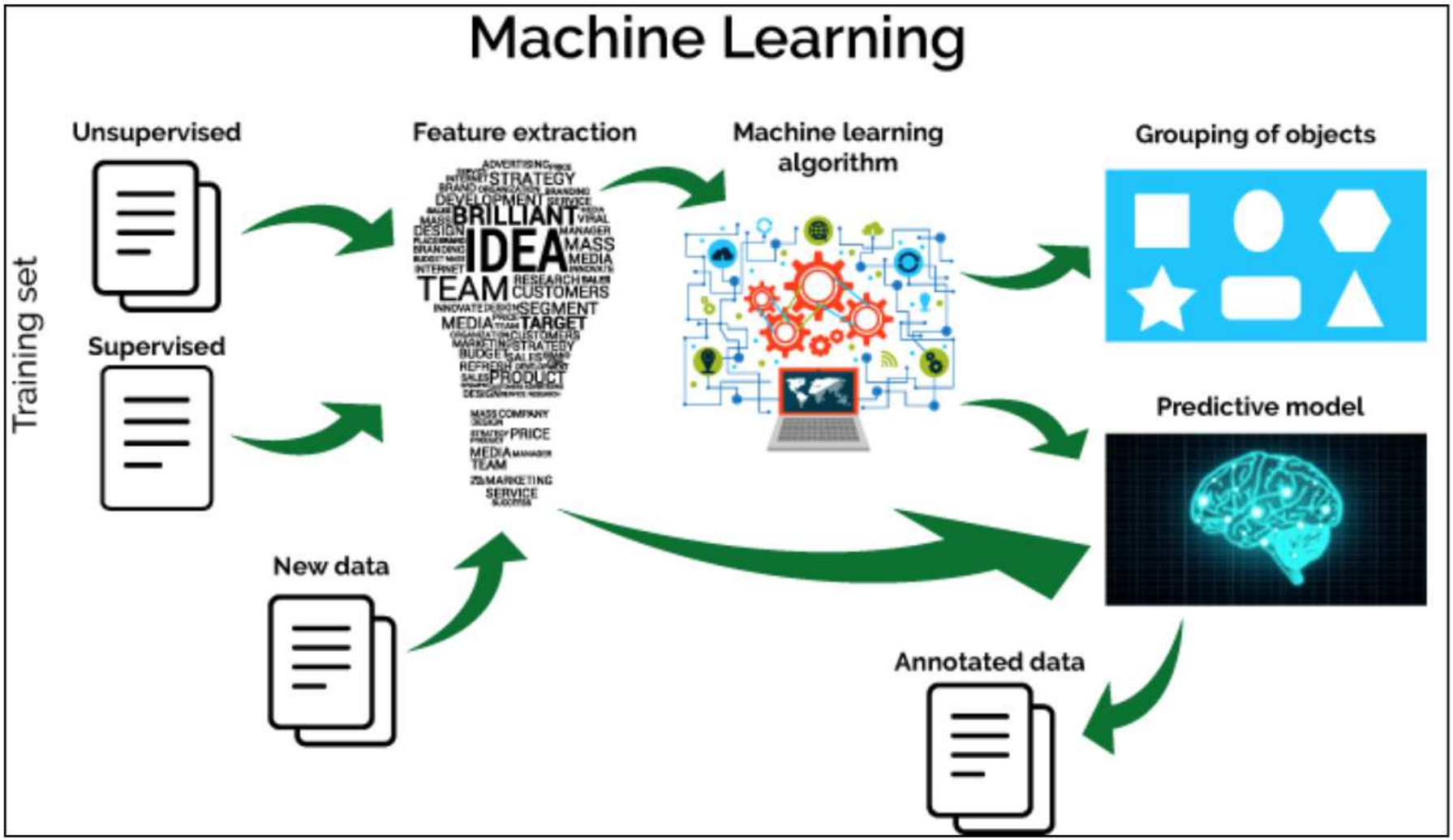
A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.

This implies that

- ML is teaching computers to generate algorithms using data without programming them explicitly.
- ML transforms data into actionable knowledge.

Machine Learning

Feed large sample data sets to find relevant mapping from inputs to outputs or find patterns



Three Types of Machine Learning

ML is divided mainly into three categories:

1. **Supervised Learning**
   - The machine is presented with inputs and the outputs corresponding to those inputs.
   - The machine learns from these inputs and applies this learning in further unseen data to generate outputs.
2. **Unsupervised Learning** doesn't have the required outputs; therefore it is up to the machine to learn and find patterns that were previously unseen.
3. **Reinforcement Learning**: the machine continuously interacts with the environment and learns through this process. This includes a feedback loop.

Good Books

- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, 2016, by Trevor Hastie, Robert Tibshirani, Jerome Friedman
- *Machine Learning: A Probabilistic Perspective*, 1st Edition, 2012, by Kevin P. Murphy.
- *Pattern Recognition and Machine Learning*, 2011, by Christopher M. Bishop.
- *Deep Learning*, 2016, by Ian Goodfellow, Yoshua Bengio, Aaron Courville