Machine Learning for the Social Sciences:
Introduction and Main Paradigms

Gian Maria Campedelli, PhD
gianmaria.campedelli@unitn.it
Department of Sociology and Social Research - University of Trento

November 11, 2020

Contents

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Duration: 6 hours (November 10th, November 17th, November 24th)

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Duration: 6 hours (November 10th, November 17th, November 24th)

Requirements: no formal requirements

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Duration: 6 hours (November 10th, November 17th, November 24th)

Requirements: no formal requirements

Test: none

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Duration: 6 hours (November 10th, November 17th, November 24th)

Requirements: no formal requirements

Test: none

Laboratories: no practical laboratories (i.e. programming). *However*, if there is a sufficient number of people interested, we can try to organize a Python seminar in the second semester

Course Outline & Info

Overview: The course seeks to provide participants with the basic ingredients to start navigating the complexity of ML, with a specific focus on social science applications

Duration: 6 hours (November 10th, November 17th, November 24th)

Requirements: no formal requirements

Test: none

Laboratories: no practical laboratories (i.e. programming). *However*, if there is a sufficient number of people interested, we can try to organize a Python seminar in the second semester

My aim: fostering the curiosity of the agnostics ($\sim$ 17.7% of the total), showcasing the potential and the challenges of this field to everyone

Tedious Bio

Hello everyone!
By now, you have learned that I am **Gian Maria Campedelli**, postdoctoral research fellow in *Computational Criminology* here at the University of Trento.

- I obtained a PhD in *Criminology* from Universita Cattolica in Milan
- From 2016 to 2019 I worked as a research associate at **Transcrime**, the *the joint research centre on transnational crime* of Università Cattolica, University of Bologna and University of Perugia
- In 2018, I also held a visiting research scholar position at **Carnegie Mellon University - School of Computer Science**, within the *Institute for Software Research*

Research Interests

My research intersects the study of crime and terrorism with computational sciences.

Specific criminological focus: **organized crime** and **terrorism**

While at Transcrime, I worked on the Horizon2020 project **PROTON**.

While at CMU, I worked on the integration of **complex networks** and **deep learning** for terrorist target forecasting.

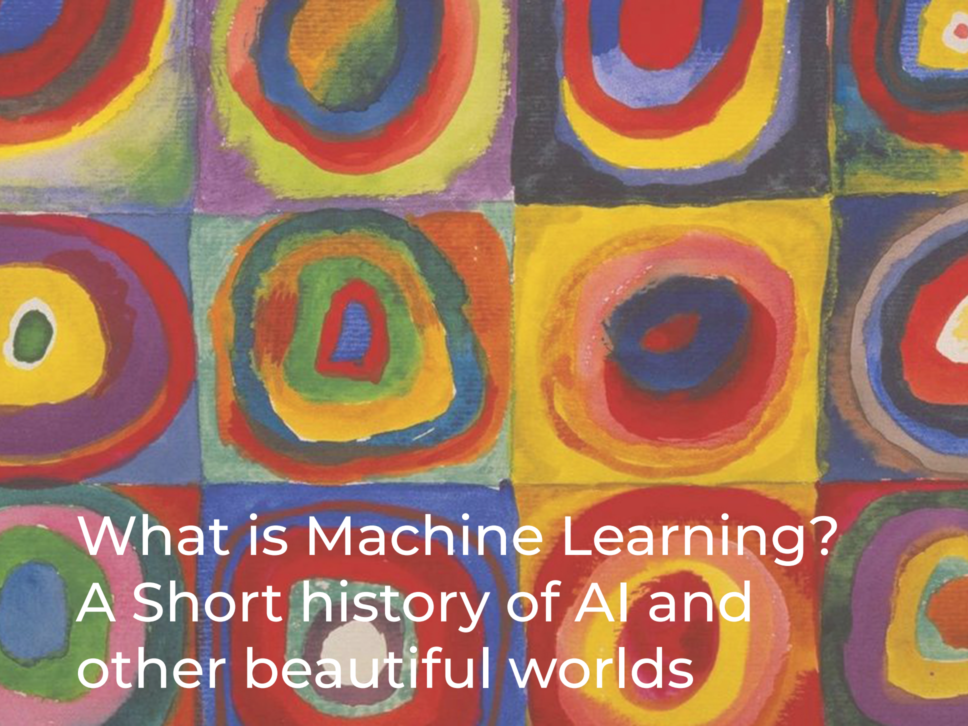
Research Interests

...besides these two main tracks:

- *Impact of COVID-19 containment policies on crime*
- *Developmental trajectories of mafia members*
- *Unsupervised learning approaches to represent terrorism dynamics from complex nets*
- *Multiple Homicide Offenders (aka Serial Killers) career modeling*
- *Pathways to increase positive societal payoffs of AI research on crime*

Now, the most important thing:

Who are you?


What is Machine Learning?
A Short history of AI and
other beautiful worlds

The Cradle of AI: Philosophy/1

The abstract idea of AI begins before modern computers. It even begins before computers and calculators (for a brief but detailed history of AI see Russell and Norvig 2020)

Aristotle first reasoned extensively on the laws that govern the rational human mind.

Following ⇒ **Hobbes**: *reasoning is a form of computation*

Throughout centuries **Leonardo, Shickard** and **Pascal** first attempted to build automatic calculators.

"The arithmetical machine produces effects which approach nearer to thought than all the actions of animals. But it does nothing which would enable us to attribute will to it, as to the animals" Blaise Pascal, Pensées (1669)

The Cradle of AI: Math and Logic/1

The mathematical study of formal logic begins with**Boole** ⇒ *the study of algorithms for logic deduction*

Godel: *incompleteness theorems* ⇒ there exist functions and propositions considered true that, however, cannot be verified (i.e., computed)

This pushes **Alan Turing** to exactly understand what functions could be computed. ⇒ *Turing Machine*: it should have been able to compute every computable function.

The study of **computation** opens the debate on the concept of **tractability**: how fast does the time of execution of a program increase as instances increase? ⇒ *Theory of NP-Completeness*

The Cradle of AI: Math and Logic/2

Besides logic and computation, **probability** is the third main area of mathematics that contributed more to the development of the origins of AI.

Today, practically all AI-designed algorithms have a *probabilistic* structure

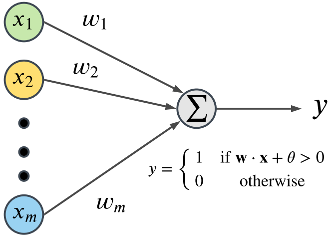
The works of **Bernoulli** (1654-1705), **Laplace** (1749-1827), and **Bayes** (1702-1761) represent the basis of the functioning of this algorithms.

The Cradle of AI: Neuroscience

Neuroscience has been a fundamental discipline for the development of AI.

Rashevsky was the first to try to apply mathematical models to the study of the nervous system. Following: **McCulloch and Pitts 1943** $\Rightarrow$ (*A Logical Calculus of Ideas Immanent in Nervous Activity*) and **Hebb 1949** (*Hebbian Learning*)

Rosenblatt 1958 $\Rightarrow$ *Perceptron*



Dartmouth College 1956

"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon

The ten formal founders of Artificial Intelligence meet in Dartmouth:

- **John McCarthy** (Dartmouth College)
- **Marvin Minsky** (Harvard University)
- **Claude Shannon** (MIT)
- **Nathaniel Rochester** (IBM)
- **Trenchard More** (Princeton University)
- **Arthur Samuel** (IBM)
- **Ray Solomonoff** (IBM)
- **Oliver Selfridge** (MIT)
- **Allen Newell** (Carnegie Tech)
- **Herbert Simon** (Carnegie Tech)

First Results...

The 1956 meeting **did not lead** to any major breakthrough but it allowed the fathers of the discipline to meet and start collaborating together.

After 1956, we observe the first **very limited** successes in the small but growing community of scientists. Programs were limited to:

- *Imitation of protocols used by humans in the resolution of problems (e.g., General Problem Solver)*
- *Proof of mathematical theorems (e.g., Geometry Theorem Prover)*
- *Development of specific programming languages for AI problems (e.g., LISP)*

...before Winter(s)

In light of the many unmet promises of AI, the field enters in the first so-called **AI winter** (1974-1980)

Following ⇒ the rise of **Expert Systems** produces new enthusiasm (medical applications, e.g., *MYCIN*) ⇒ New success fosters industry attention and attracts financial capitals

Disillusion and disappointment of investors led to the **second AI winter** (1987-1993)

The Rebirth

After two periods of stagnation, AI undergoes a new acceleration due to three main developments:

- Establishment of the **connectivist approach** (complementary to the symbolist one) $\Rightarrow$ creation of new *neural networks*
- Approach to problems governed by the **scientific method**
- Development of faster and more powerful **computers/hardware** + increasing availability of **data**

AI strikes back

The combination of these factors facilitated many of the results predicted by the fathers of AI decades before.

- 1997: **IBM DeepBlue** wins a chess game against Garry Kasparov
- 2001: **NASA** develops a software to monitor and plan on-board operations on a spaceship
- 2005: **Stanford** wins the DARPA Gran Challenge
- 2007: **CMU** wins the DARPA Urban Challenge

Recent Breakthroughs

- 2011: **IBM Watson** beats two of the longest-running champions of *Jeopardy!* (reaching a 1M USD prize)
- 2016: **AlphaGo (Google+DeepMind)** wins against GO's European champion Fan Hui (5-0) and world-level champion Lee Sedol (4-1)
- 2017: **Libratus (CMU)** gains 1.7M USD playing 120,000 Texas Hold'em hands. For the first time a machine proofs to be able to win against human champions in a game with no *perfect information symmetry*

A First Dichotomy

The progress (and the failures) of the last 50 years led to a new conceptualization of AI:

- **Weak (or applied) AI**: a program (or a set of programs) that can solve well-defined, limited tasks
- **Strong (or general) AI**: a sentient machine with mind and consciousness that can act intelligently in any kind of realm

A First Dichotomy

The progress (and the failures) of the last 50 years led to a new conceptualization of AI:

- **Weak (or applied) AI**: a program (or a set of programs) that can solve well-defined, limited tasks

So, what is Machine Learning?

Within weak-applied AI, stands **Machine Learning (ML)**, which aims at enabling machines to learn a certain task without being explicitly programmed to do so.

Mitchell's Definition

Mitchell 1997: A machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system reliably improves its performance P at task T, following experience E

Contextualizing Mitchell's Definition

- **A machine** $\rightarrow$ *An algorithm (or a set of algorithms)*
- **A task** $\rightarrow$ *Predicting whether an item belongs to class* A *or* B
- **Experience** $\rightarrow$ *Data inputs*
- **Performance metric** $\rightarrow$ *Predictive accuracy, F1-score, etc.*

ML in our lives

Today, most of the technological instruments that we use every day rely on ML or similar forms (e.g., Deep Learning) to function.

- **Netflix** uses ML-empowered recommendation systems to suggest movies we may like
- **Smartphones'** text autocomplete options run on specific neural networks
- **Virtual Personal Assistants** as *Amazon Alexa* revolve around speech recognition and NLP
- **Social media** platforms exploit algorithms to show certain contents to us
- *And many many more...*


A Learning Bipolarism:
Supervised and
Unsupervised Learning

Bipolarism? Let's pretend so...

ML entails a complex array of different learning paradigms.

The most common dichotomy divides between:

1. **Supervised Learning**
2. **Unsupervised Learning**

Among the other paradigms are *Reinforcement Learning*, *Semi-supervised Learning*, *Self-supervised learning* $\Rightarrow$ not covering them in this course

Supervised Learning: Intro

Supervised Learning refers to all the classes of problems for which we have a known target variable (e.g., predicting whether a picture contains a dog or a cat).

A more formal definition

We first assume we are given N examples $\mathbf{x}_n \in \mathbb{R}^D$ and N corresponding scalar labels $\mathbf{y}_n \in \mathbb{R}$. In the supervised learning setting we obtain the pairs $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N)$. From these pairs, we would like to estimate a predictor $f(\cdot, \theta) : \mathbb{R}^D \to \mathbb{R}$ The aim is to find a good parameter $\theta*$ such that:

$$f(\mathbf{x}_n, \theta^*) \approx y_N \;\; \forall n = 1, ..., N$$

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1			
2			
3			

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	
2			
3			

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	Regression
2			
3			

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	Regression
2	Clinical History, Age	Status (Dead/Alive)	
3			

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	Regression
2	Clinical History, Age	Status (Dead/Alive)	Classification
3			

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	Regression
2	Clinical History, Age	Status (Dead/Alive)	Classification
3	Color, Weight, Height	Species (Reptile, Fish, Bird)	

Supervised Learning: Tasks

The fundamental distinction between tasks in Supervised Learning refers to:

- **Classification**: $\mathbf{y}$ is discrete
- **Regression**: $\mathbf{y}$ is continuous

Case	$\mathbf{x}_n$	$\mathbf{y}_n$	Task?
1	Age, Education, Sex	USD Net Income	Regression
2	Clinical History, Age	Status (Dead/Alive)	Classification
3	Color, Weight, Height	Species (Reptile, Fish, Bird)	Classification

Supervised Learning: Classification

Classification encompasses a number of different typologies:

Classification encompasses a number of different typologies:

1. **Binary classification**: a task with only two class labels (e.g., dog/cat, awarded/not awarded...)

Supervised Learning: Classification

Classification encompasses a number of different typologies:

1. **Binary classification**: a task with only two class labels (e.g., dog/cat, awarded/not awarded...)
2. **Multi-class Classification**: a task with $L > 2$ labels (e.g., digit recognition, face classification)

Supervised Learning: Classification

Classification encompasses a number of different typologies:

1. **Binary classification**: a task with only two class labels (e.g., dog/cat, awarded/not awarded...)
2. **Multi-class Classification**: a task with $L > 2$ labels (e.g., digit recognition, face classification)
3. **Multi-label Classification**: a task where multiple labels can be assigned to an instance (e.g., object recognition)

Supervised Learning: Algorithms

- **Classification Tasks:** *decision trees, random forests, boosting, support vector machines, neural networks, nearest neighbor algorithms, Naive Bayes, bagging...*
- **Regression Tasks:** *logistic regression, linear regression, polynomial regression, Spline regression, penalized regression...*

Supervised Learning: Algorithms

- **Classification Tasks:** *decision trees, random forests, boosting, support vector machines, neural networks, nearest neighbor algorithms, Naive Bayes, bagging...*
- **Regression Tasks:** *logistic regression, linear regression, polynomial regression, Spline regression, penalized regression...*

$\Rightarrow$ We will see some of these approaches in the next lecture

Unsupervised Learning: Intro

Unsupervised Learning concerns the classes of problem in which there is not outcome/target measure, and the goal is therefore to find and describe patterns and associations among inputs.

A more formal definition

Contrarily to what happens in supervised learning, unsupervised learning problems do not experience any *teacher* signal. The machine only receives inputs $\mathbf{x}_n \in \mathbb{R}^D$ that form a random p-vector X having joint density $\Pr(X)$

Unsupervised Learning: Tasks

Unsupervised Learning cover many different tasks that are often common in data science applications, including:

- **Clustering**: Finding communities/subgroups of objects that are similar according to a certain criterion (generally *intra-cluster similarity* and *inter-cluster dissimilarity*). Different families of clustering approaches (e.g., distance-based, density-based, hierarchical...)

Unsupervised Learning: Tasks

Unsupervised Learning cover many different tasks that are often common in data science applications, including:

- **Clustering**: Finding communities/subgroups of objects that are similar according to a certain criterion (generally *intra-cluster similarity* and *inter-cluster dissimilarity*). Different families of clustering approaches (e.g., distance-based, density-based, hierarchical...)
- **Anomaly Detection**: Detecting instances that significantly deviate from a learned distribution of previous instances

Unsupervised Learning: Tasks

Unsupervised Learning cover many different tasks that are often common in data science applications, including:

- **Clustering**: Finding communities/subgroups of objects that are similar according to a certain criterion (generally *intra-cluster similarity* and *inter-cluster dissimilarity*). Different families of clustering approaches (e.g., distance-based, density-based, hierarchical...)
- **Anomaly Detection**: Detecting instances that significantly deviate from a learned distribution of previous instances
- **Dimensionality Reduction**: Representing X in a lower-dimension feature vector while preserving key properties of the data (overcoming the *curse of dimensionality* reduction

Unsupervised Learning: Tasks

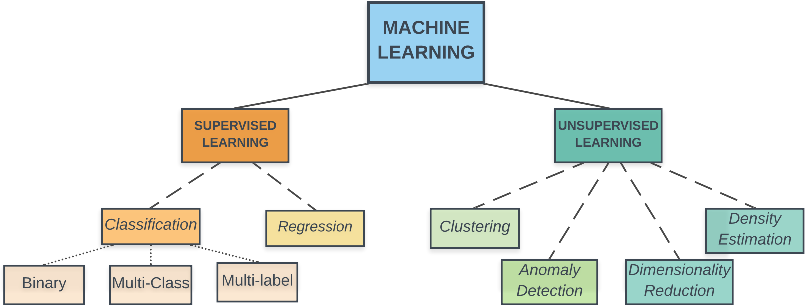
Unsupervised Learning cover many different tasks that are often common in data science applications, including:

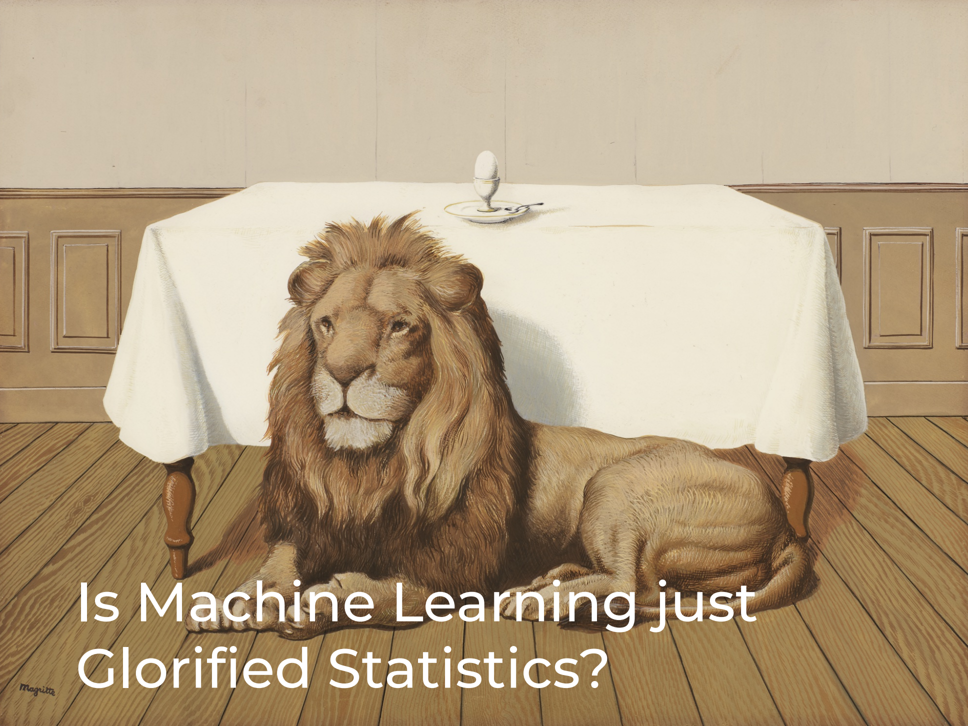
- **Clustering**: Finding communities/subgroups of objects that are similar according to a certain criterion (generally *intra-cluster similarity* and *inter-cluster dissimilarity*). Different families of clustering approaches (e.g., distance-based, density-based, hierarchical...)
- **Anomaly Detection**: Detecting instances that significantly deviate from a learned distribution of previous instances
- **Dimensionality Reduction**: Representing X in a lower-dimension feature vector while preserving key properties of the data (overcoming the *curse of dimensionality* reduction
- **Density Estimation**: Constructing an estimate of the probability density function using a set of data points

Unsupervised Learning: Algorithms

- **Clustering**: *k-means, DBSCAN, Spectral Clustering, OPTICS, Ward Hierarchical...*
- **Anomaly Detection**: *One-class SVM, Isolation Forest, Local Outlier Factor...*
- **Dimensionality Reduction**: *Singular Value Decomposition, Principal Component Analysis, Linear Discriminant Analysis...*
- **Density Estimation**: *Gaussian Mixtures, Kernel Density Estimation...*

Graphic Recap





Is Machine Learning just Glorified Statistics?

ML outside CS

In the last decade, ML has gained popularity outside the main fields it originated from. Three main factors behind this process:

- **Data:** explosion of data availability $\Rightarrow$ *Digitalization, social media data, etc.*
- **Software:** increasing accessibility of programs/languages to build/deploy ML algorithms/frameworks
- **Hardware:** more and more powerful CPU, cheaper access to GPU

ML challenges outside ML

The hype associated with ML sparked many debates on the actual potential of *intelligent* algorithms to solve tasks in many domains. Especially in the **social sciences**.

The **social sciences** are often characterized by *extremely* complex mechanisms.

Human behavior is **dynamic**, often **fat tailed**, often **irrational**.

Mainstream ML is often deployed on **fixed, balanced, controlled** settings.

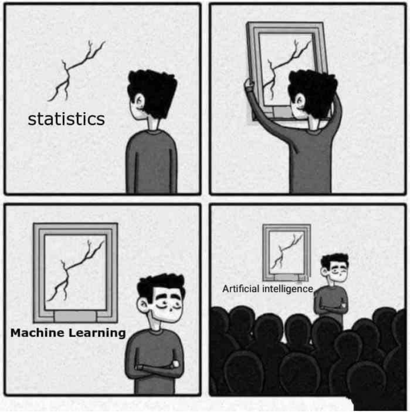
Hype vs Reality

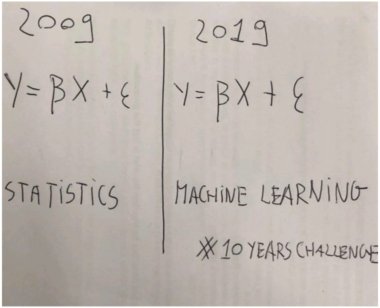
The hype on ML/DL in the last 10 years **shocked** academia and industry.

ML has become a *buzzword* to attract funding from public institutions and private companies

This also led to negative societal consequences (e.g., the use of ML in *policing* and *criminal justice*) (see, for instance, Angwing et al. 2016).

...and finally sparked vivid debates on fairness, ethics, social justice and the relationship with **Statistics**


statistics
Machine Learning
Artificial intelligence


2009
2019
$Y = \beta X + \epsilon$
$Y = \beta X + \epsilon$
STATISTICS
MACHINE LEARNING
#10 YEARS CHALLENGE

Machine Learning vs Statistics

While there is significant overlap between ML and Stats concerning certain concepts and methods (e.g., Logistic Regression), there are several differences between the two:

Machine Learning vs Statistics

While there is significant overlap between ML and Stats concerning certain concepts and methods (e.g., Logistic Regression), there are several differences between the two:

- **Data generating process:** Statistics assumes data are generated by a given stochastic model, ML does not (data generating process is considered unknown)

Machine Learning vs Statistics

While there is significant overlap between ML and Stats concerning certain concepts and methods (e.g., Logistic Regression), there are several differences between the two:

- **Data generating process:** Statistics assumes data are generated by a given stochastic model, ML does not (data generating process is considered unknown)
- **Inference vs Prediction:** Statistics is mostly concerned with inferring relationships among variables (β is crucial). ML generally focuses on predicting/forecasting patterns ($\hat{Y}$ is crucial)

Machine Learning vs Statistics

While there is significant overlap between ML and Stats concerning certain concepts and methods (e.g., Logistic Regression), there are several differences between the two:

- **Data generating process:** Statistics assumes data are generated by a given stochastic model, ML does not (data generating process is considered unknown)
- **Inference vs Prediction:** Statistics is mostly concerned with inferring relationships among variables (β is crucial). ML generally focuses on predicting/forecasting patterns ($\hat{Y}$ is crucial)
- **Validation:** Stats often does not care about validation. ML, given its aims, is fundamentally concerned with validation (e.g., T-T paradigm, cross-validation...)

Machine Learning vs Statistics

While there is significant overlap between ML and Stats concerning certain concepts and methods (e.g., Logistic Regression), there are several differences between the two:

- **Data generating process:** Statistics assumes data are generated by a given stochastic model, ML does not (data generating process is considered unknown)
- **Inference vs Prediction:** Statistics is mostly concerned with inferring relationships among variables (β is crucial). ML generally focuses on predicting/forecasting patterns ($\hat{Y}$ is crucial)
- **Validation:** Stats often does not care about validation. ML, given its aims, is fundamentally concerned with validation (e.g., T-T paradigm, cross-validation...)
- **Amount of data:** Classic Stats approaches can work well on small datasets. ML generally requires wider samples. DL even more.

ML: Beyond Forecasting

ML traditionally deals with forecasting/prediction. However ML can also be used for causal discovery. Two main ways:

- ML approaches/algorithms can be used for **theory testing/development** $\Rightarrow$ prediction as a diagnostic tool to assess validity of hypotheses
- ML algorithms expressively designed for **discovering/validating causal links** in observational studies

Pearl and Mackenzie 2018: *no true artificial intelligence without machines that can understand causality.*

Heartfelt Suggestions

Increasing number of models/methods/approaches for investigating questions and research problems in the social sciences. In the context of the *Statistics vs ML* debate, I suggest to:

Heartfelt Suggestions

Increasing number of models/methods/approaches for investigating questions and research problems in the social sciences. In the context of the *Statistics vs ML* debate, I suggest to:

- **Know your problem**: sophisticated methods won't save you from blurry specifications + *No Free Lunch Theorem* as proposed by Wolpert and Macready 1997

Heartfelt Suggestions

Increasing number of models/methods/approaches for investigating questions and research problems in the social sciences. In the context of the *Statistics vs ML* debate, I suggest to:

- **Know your problem**: sophisticated methods won't save you from blurry specifications + *No Free Lunch Theorem* as proposed by Wolpert and Macready 1997
- **Know your data**: regardless of the main analytical strategy it is *fundamental* to explore your data (data in the social sciences are *messy*)

Heartfelt Suggestions

Increasing number of models/methods/approaches for investigating questions and research problems in the social sciences. In the context of the *Statistics vs ML* debate, I suggest to:

- **Know your problem**: sophisticated methods won't save you from blurry specifications + *No Free Lunch Theorem* as proposed by Wolpert and Macready 1997
- **Know your data**: regardless of the main analytical strategy it is *fundamental* to explore your data (data in the social sciences are *messy*)
- **The importance of complementing**: especially in long projects (e.g., PhD theses), combining different approaches is critical $\Rightarrow$ *forecasting + inference*

Heartfelt Suggestions

Increasing number of models/methods/approaches for investigating questions and research problems in the social sciences. In the context of the *Statistics vs ML* debate, I suggest to:

- **Know your problem**: sophisticated methods won't save you from blurry specifications + *No Free Lunch Theorem* as proposed by Wolpert and Macready 1997
- **Know your data**: regardless of the main analytical strategy it is *fundamental* to explore your data (data in the social sciences are *messy*)
- **The importance of complementing**: especially in long projects (e.g., PhD theses), combining different approaches is critical $\Rightarrow$ *forecasting + inference*
- **Never, never, never follow the hype**: do not choose ML just because it sounds cool. A fair share of ML applications in the social sciences at the moment is *garbage*: Google Scholar won't forget

Thanks for your attention!

email: gianmaria.campedelli@unitn.it
Twitter: @CampedelliGian

Julia Angwing et al. (2016). “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks.”. In: *ProPublica*. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

D. O. Hebb (1949). *The Organization of Behavior: A Neuropsychological Theory*. Inglese. 1st Edition. Mahwah, N.J: Psychology Press. ISBN: 978-0-8058-4300-2.

Warren S. McCulloch and Walter Pitts (Dec. 1943). “A logical calculus of the ideas immanent in nervous activity”. en. In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259 (visited on 11/09/2020).

Tom M. Mitchell (1997). *Machine Learning by Tom M. Mitchell (1997) Hardcover*. McGraw-Hill Science/Engineering/Math.

Judea Pearl and Dana Mackenzie (2018). “AI can’t reason why”. In: *Wall Street Journal*.

Bibliography II

F. Rosenblatt (1958). “The perceptron: A probabilistic model for information storage and organization in the brain”. In: *Psychological Review* 65.6. Place: US Publisher: American Psychological Association, pp. 386–408. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0042519.

Stuart J. Russell and Peter Norvig (Feb. 2020). *Artificial Intelligence: A Modern Approach*. Inglese. 4 edizione. Hoboken: Pearson College Div. ISBN: 978-0-13-461099-3.

D. H. Wolpert and W. G. Macready (Apr. 1997). “No free lunch theorems for optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1. Conference Name: IEEE Transactions on Evolutionary Computation, pp. 67–82. ISSN: 1941-0026. DOI: 10.1109/4235.585893.

Artwork Reference

Title Slide: *Peinture (Etoile Bleue), Joan Miro (1927)*

What is Machine Learning? A short history of AI and other beautiful worlds: *Farbstudie - Quadrate und konzentrische Ringe, Vasilij Vasil'evič Kandinskij (1913)*

A Learning Bipolarism: Supervised and Unsupervised Learning: *Archeological Reminiscence of Millet's Angelus, Salvador Dalì (1934)*

Is Machine Learning just glorified Statistics? *Le Repas De Noces, René Magritte (1940)*

Closing Slide: *The Persistence of Memory, Salvador Dalì (1931)*