PhD Programme in Sociology and Social Research
University of Trento

Machine Learning for the Social Sciences

Gian Maria Campedelli, PhD

September 2020

Course Overview

This short course (denominated ML4SS) aims at providing doctoral students with the necessary elementary ingredients to familiarize with machine learning for the investigation of social science research problems. It will be delivered in three distinct lessons, each lasting two hours. The contents of ML4SS assume that the participants have at least an intermediate understanding of statistics (or quantitative methods in general), but the lessons are designed to be enjoyable and open for any audience, including students belonging to qualitative backgrounds. Given the increasing diffusion of machine learning and artificial intelligence-related applications in the real world, a basic understanding of what learning algorithms can (and cannot) do in the social science landscape is fundamental to unfold future scientific developments — and anticipate possible societal risks.

Dates

- **Lesson 1:** November 10th, 2020 (10:00-12:00)
- **Lesson 2:** November 17th, 2020 (10:00-12:00)
- **Lesson 3:** November 24th, 2020 (10:00-12:00)

Contents

Lesson 1: Introduction and Main Paragidms

Outline: The first lesson will cover some introductory topics:

- *What is Machine Learning? A short history of Artificial Intelligence and other beautiful worlds*
- *Is Machine Learning just glorified Statistics?*
- *A Learning Bipolarism: Supervised Learning and Unsupervised Learning*
- *Important Concepts: Training, Out-of-sample Performance, Over-fitting and Regularization*

Lesson 2: Algorithms and Applications

Outline: After having explained the difference between supervised and unsupervised learning, the second lesson will delve into the most common algorithms in the two areas:[1]

[1] Minor changes may apply.

- *Supervised Learning: Classification and Regression*
- *Supervised Learning Methods: Tree-based Algorithms (Decision Trees, Random Forests, Gradient Boosting), Regression, Shrinkage Methods (e.g., Ridge and Lasso Regression), Bayesian Methods, Main Neural Networks Architectures*
- *Unsupervised Learning Methods: Dimensionality Reduction, Clustering and Representation Learning*

Lesson 3: Advanced Topics and New Directions

Outline: The last lesson will be divided into two main tracks. During the first part, I will conclude the overview on technical applications for the social sciences with a specific focus on models for temporal and spatial data and machine learning for networks. The second part, instead, will be dedicated to discuss the societal consequences of biased algorithms.

- *Learning Over Time: Recurrent Neural Networks, Survival Methods and Point Processes*
- *Learning Across Space: Convolutional Neural Networks, GIS and Machine Learning*
- *Connecting the dots: Graph Learning*
- *Ethical and political implications of algorithmic decision-making processes: the case of criminal justice and predictive policing*

Suggested readings:

Lesson 1

Readings for all:

- Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199–231. https://doi.org/10.1214/ss/1009213726
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Alstyne, M. V. (2009). Computational Social Science. *Science, 323*(5915), 721–723. https://doi.org/10.1126/science.1167742
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, LIX* (236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Lesson 2

Dedicated reading: sociologists

- Molina, M., Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology, 45*(1), 27–45. https://doi.org/10.1146/annurev-soc-073117-041106

Dedicated reading: economists

- Athey, S. (2018). The Impact of Machine Learning on Economics. *The Economics of Artificial Intelligence: An Agenda*, 507–547. https://www.nber.org/chapters/c14009

Dedicated reading: political scientists

- Cranmer, S. J., Desmarais, B. A. (2017). What Can We Learn from Predictive Modeling? *Political Analysis, 25*(2), 145–166. https://doi.org/10.1017/pan.2017.3

Dedicated reading: criminologists

- Brennan, T., Oliver, W. L. (2013). The Emergence of Machine Learning Techniques in Criminology. *Criminology & Public Policy, 12*(3), 551–562. https://doi.org/10.1111/1745-9133.12055

Lesson 3

Readings for all:

- Angwing, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Barabas, C., Virza, M., Dinakar, K., Ito, J., Zittrain, J. (2018). Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. *ACM Conference on Fairness, Accountability and Transparency*, 62–76. http://proceedings.mlr.press/v81/barabas18a.html
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894