**Syllabus**

**Machine Learning for Natural Language Processing**

* **Instructor:**  Arieda Muço (MucoA@ceu.edu)
* **Credits:** 4 ECTS
* **Term:** Winter 2021
* **Course level:** Ph.D./MA Economics
* **Prerequisites**: Math pre-session, Effective Computing for Economics, Econometrics

**Course description**

We will cover methods from the machine learning literature that we view as an important toolset for empirical economics. We will discuss supervised learning methods for regression and classification, unsupervised learning methods, as well as text-analysis applications. Throughout the course we highlight the intersection of ML and econometrics. We will use Python for algorithm implementation.

**Learning outcomes**

* Understand and use Machine learning techniques broadly and in economic applications.
* Applying Natural Language Processing (NLP) techniques to Machine Learning.
* Validation and Cross-validation.
* Overfitting, Regularization, and Tuning Parameters.
* Natural Language Toolkit (NLTK)

**Learning activities and teaching methods**

This is a mixture between lecture type and participatory course. Students participate in live coding together with the instructors. This format facilitates quick and frequent feedback (from instructors and peers) and enables students to achieve mastery in ML and computing skills.

**Assessment (including minimum pass requirement and grading)**

● Active class participation is required. (20% of grade)

● Take-home assignments solving computational tasks. (80% of grade)

**Technical/laptop requirement**

Students must bring a laptop with a Mac, Linux, or Windows operating system (not a tablet, Chromebook, etc.) that they have administrative privileges. Installation instructions and assistance available for necessary software packages. At the beginning of the course students will receive an invitation to join the course Slack workspace.

**Course schedule and materials for each session**

1. Linear Regression and Classification (Logistic Regression)

2. Linear Model Selection and Regularization (Lasso, Ridge, Elastic-Nets)

3. Nonlinear methods (Support Vector Machines, Tree Based Methods)

4. Unsupervised Learning: Cluster Analysis, Principal Components

5. Natural Language Toolkit, Bag of Words, Term-Frequency Inverse-Document Frequency

6. Introduction to Natural Language Processing Techniques (Naive Bayes, Sentiment Analysis)

**Reading list**

Bengfort, Benjamin, Rebecca Bilbro, and Tony Ojeda, Applied Text Analysis with Python: Enabling Language Aware Data Products with Machine Learning

Békés, Gábor, and Gábor Kézdi. Data Analysis for Business, Economics, and Policy. Cambridge University Press, 2021

Dan Jurafsky and James H. Martin, Speech and Language Processing

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, The Elements of Statistical Learning

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning

Sarah Guido, and Andreas Müller, Introduction to Machine Learning with Python: A Guide for Data Scientists