Syllabus

Course title	Machine Learning Concepts
Instructor	Robert Lieli, professor Janos Divenyi, practicing data scientist
Email	lielir@ceu.edu / divenyij@ceu.edu
Office hours	by appointment
Credits	1 US credit (2 ECTS credits)
Term	Winter 2024
Course level	MS
Prerequisites	DA1, DA2, DA3
Course drop	per university/department policy

1. COURSE DESCRIPTION

Content

The course introduces the basic ideas underlying the group of statistical estimators often referred to as "machine learning" (ML) methods. These methods are particularly suitable for solving prediction and classification tasks with many predictors in a flexible way. Though the course focuses on theory, there will be seminars discussing implementation in R along with some applications.

Relevance

It's hardly necessary to justify why the ability to work with data is a critical skill in this day and age. So please insert your own *bon mot* about data science or big data here.

2. LEARNING OUTCOMES

Key outcomes. By the end of the course, students will be able to

- Understand what "regularization" means and how it induces a bias-variance tradeoff fundamental to ML methods
- Design simulation exercises to understand the statistical properties of ML estimators
- Use OLS regression and various penalized regression methods (lasso, ridge) for prediction
- Understand the ideas behind principal component analysis
- Implement ML estimators in R

Other outcomes. The course will also help develop skills in the following areas.

Learning Area	Learning Outcome
Critical Thinking	Understanding what machine learning methods can and cannot do and what their basic principles are
Quantitative Reasoning	Using machine learning methods to analyze data

Technology Skills	Familiarity with R
Interpersonal Communication Skills	The ability to use and present data as evidence is a critical communication skill
Management Knowledge and Skills	
Cultural Sensitivity and Diversity	
Ethics and Social Responsibility	

3. READING LIST

Textbook

- [ISL] James, G., D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning*. Springer. A very accessible introduction to machine learning methods. Freely available online.

Article

- Mullainathan, S. and J. Spiess (2017): "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 31, pp. 87-106.

4. TEACHING METHOD AND LEARNING ACTIVITIES

The course will involve a mix of lectures and seminars. The lectures cover the theory while the seminars are focused on implementation and applications.

5. ASSESSMENT

This is a short course; there will be one take-home assignment at the end of the course. Grading will be in line with CEU's grading guidelines.

6. TECHNICAL REQUIREMENTS

We will post all assignments, course materials, announcements, etc., on the CeuLearning website. It is your responsibility to check this site regularly. Students need to bring their own laptop to the seminars with python3.10 installed (the same environment you used for DA3).

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
1	• Introduction to statistical learning: basic setup and concepts. Assessing model accuracy: in-sample and out-of-sample goodness of fit. • The bias-variance tradeoff. Underfit and overfit. • Using computer simulations to study the properties of estimators. • Seminar: implementation in Python	ISL 2.1, 2.2.1-2.2.2,
2	• From linear regression to lasso regression. Variable selection. • Regularization as a general idea. Regularization and the bias-variance tradeoff. • Seminar: implementation in Python	ISL 3.1-3.2, ISL 6.2, article
3	• A more in-depth look at how variable selection with lasso works. Other types of regularized regressions (ridge). • (Time permitting: unsupervised learning and principal components) • Seminar: implementation in Python • Assignment given	ISL 3.1-3.2, ISL 6.2, article (ISL 6.3, 10.1-10.2)

8. SHORT BIO OF THE INSTRUCTORS

Robert Lieli received his Ph.D. in economics from the University of California, San Diego in 2004. Before joining CEU full time, he worked at the University of Texas at Austin and the National Bank of Hungary. His research area is econometrics with particular emphasis on binary prediction, forecasting, and treatment effects. He has published in top economics journals such as the Journal of Econometrics, the Journal of Business and Economic Statistics and the Journal of the European Economic Association.

János Divényi earned his Ph.D. in economics in 2020 at the Central European University. He has been working as a data scientist for more than 10 years, currently focusing on the analytical capabilities of a marketing cloud platform.