Doctoral School in Finance and Economics

University of Luxembourg 2020-2021

Course ID

Doctoral Seminar on Machine Learning in Economics, Management, and Finance

1. Course details

Semesters:	1
Credit rating:	1 ECTS
Teaching units:	15
Pre-requisite(s):	The prerequisites for this course are introductory courses in statistics or probability, prior programming experience in **R** and/or the willingness to learn **R,** and a good quantitative background.
Lecturers:	Dr. Rutger W. Poldermans (rutger.poldermans@uni.lu)
Administrator:	Roswitha Glorieux (roswitha.glorieux@uni.lu)
Tutors:	None
Seminar times and rooms:	See plan of semester B-28
Tutorial times and rooms:	None
Communications	**It is important that students should regularly read their University e-mails, as important information will normally be communicated this way.**
Mode of assessment:	Class Participation & Assignments
Examination Periods:	None
Course WebPage:	Moodle.uni.lu

2. Aims and objectives

Aims

This is doctoral level course covers the basic theory of machine learning and its relevance for empirical researchers in economics and finance. Topics being studied include optimal prediction rules, cross-validation, regularization (lasso and ridge), classification, nonlinear models, tree-based methods and unsupervised learning. Computing is done in the statistical programming language **R**, through tutorial sessions and two assignments.

Learning Objectives

Upon successful completion of this course students will be able to:

- recognize the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
- explain the strengths and weaknesses of many popular machine learning approaches.
- implement various algorithms in a range of real-world applications with **R**

3. Plan of semester

Dates	Room	Time	Seminar Lecture	Topic	Deadlines
Apr. 6	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-
Apr. 8	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-
Apr. 13	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-
Apr. 15	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-
Apr. 20	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-
Apr. 22	B-28	14:00 – 16:30	Lecture + Tutorial	TBA	-

4. Course details (by topics)

See attached syllabus

5. Further information about assessment

Examination(s) There is no exam. The final grade (Pass/Fail) is based on class participation and two assignments.

Weighting: See attached syllabus

Date: See plan of semester

MACHINE LEARNING METHODS IN ECONOMICS, MANAGEMENT, AND FINANCE

Dr. Rutger W. Poldermans
University of Luxembourg

BRIEF DESCRIPTION

This doctoral level course covers the basic theory of machine learning and its relevance for empirical researchers in economics, management and finance.

Topics studied include optimal prediction rules, linear regression, cross-validation, regularization (lasso and ridge), classification, tree-based methods, support vector machines and unsupervised learning. Computing is done in the statistical programming language **R**, through tutorial sessions and two assignments.

The main goal of this course is to facilitate a learning environment in which the students are able to master "state-of-the-art" methods in machine learning and apply these in the context of their individual research interests.

INTENDED AUDIENCE

This course is intended for PhD students in Economics, Management, and Finance. However, all PhD students with a natural interest in quantitative methods are encouraged to participate.

TEXT AND MATERIALS

The required study material for this course is

- An Introduction to Statistical Learning with applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
- Software (open source)
    - The R project for statistical computing (https://www.r-project.org/)
    - RStudio (https://rstudio.com/)
- Other course material, e.g. lecture notes, additional data sets or computer exercises, will be made available on the Moodle site of this course

Students interested in additional reading may consider the following references:

- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Berlin: Springer
- Hastie T., Tibshirani R., Wainwright, M., 2015. Statistical Learning with Sparsity: The Lasso and Generalization.
- Burkov, A., 2019. The Hundred-Page Machine Learning Book. Quebec City, Can.: Andriy Burkov.
- Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. The MIT Press.
- Van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge, UK: Cambridge Univ. Pres.
- Athey, S., Imbens, G.W., 2019, "Machine Learning Methods That Economists Should Know About," Annual Review of Economics 11, 685-725
- Hastie, T., Tibshirani R., Tibshirani, R.J., 2017. Extended comparisons of best subset selection, forward stepwise selection, and the lasso. arXiv:1707.08692 [stat.ME]
- Friedberg, R., Tibshirani, J., Athey, S., Wager, S., 2020. Local linear forests. arXiv:1807.11408 [stat.ML]
- Belloni, A., Chernozhukov, V., Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. J. Econ. Perspect. 28:29–50

GRADING

- Regular class participation: 10% [Every class meeting]
  Students need to come to the classroom with their minds on the material being discussed. Just "showing up" for class does not count for class participation.
- Assignment 1: 40% [Due: April 15]
  The first assignment consists of a number of simulated data sets based on so-called data-generating processes (DGPs) with specific characteristics being given only partially. The main goal is to select a machine learning technique for linear models that provides the best predictions in terms of the mean squared error (MSE) for each data set individually.
- Assignment 2: 50% [Due: April 30]
  The second assignment concerns different classification data sets. Each student is expected to conduct a profound analysis on a data set of personal interest and write a concise report with a maximum of ten A4 pages. The report must include a clear description of the data and the task for the project, a sensible comparison of different classification methods, well-organized results and a solid conclusion with future recommendations.

LECTURE SCHEDULE

(Note: Tentative and subject to change)

Date	Subject
Apr. 6	Bias-variance tradeoff and resampling methods. Reading: 2.1-2.3, 5.1.1- 5.1.4, 5.2
Apr. 8	(High dimensional) linear regression and regularization. Reading: Ch. 3 and Ch. 6. First assignment avail.
Apr. 13	Classification. Reading: Ch. 4, 5.1.5,
Apr. 15	Tree-based methods. Reading: Ch. 8. Deadline first assignment
Apr. 20	Support vector machines. Reading: Ch. 9. Second assignment avail.
Apr. 22	Unsupervised Learning. Reading: Ch. 10
Apr. 30	Deadline second assignment.