

Universität Hamburg

Machine Learning for Economics and Finance

Ole Wilms

SoSe 2023

1. Contents

This bachelor level course will give an introduction to machine learning techniques with a particular focus on how they can be applied in economics and finance. We will review the standard techniques used in machine learning and learn how to apply them in practice. A large focus will be on choosing appropriate methods to prevent the models from overfitting the data. The course will be mainly about supervised learning methods that can be used to make predictions. However, if time permits, we might also cover the basics of unsupervised learning problems.

2. Learning goals

This course is an introduction to machine learning in R with applications to finance. Upon completion of this course, you should be able to:

- (1) Use the free software R to solve key tasks in machine learning including loading and cleaning big datasets, using libraries/r-packages, summarize and visualize data
- (2) Apply supervised learning methods to solve prediction problems in economics and finance
- (3) Analyze the main benefits and limitations of the supervised learning methods we cover in class
- (4) Evaluate and compare the performance of different methods

3. Covered topics

The course puts a special emphasis on the **application** of machine learning techniques. For this, you will learn how to practically analyze large-scale economic data and make predictions. In particular, we focus on the following topics throughout the course:

- Introduction to statistical computing. Students will learn the basics about using computers to analyze big data, with a special emphasis on R, and the most common machine learning libraries in R.
- Working with data. Data is rarely found in a perfectly usable form. You will learn how to clean the data to make it usable.
- Supervised learning techniques have become very advanced. We cover the basic methods including
 - linear models such as Lasso and Ridge regressions
 - decision trees
 - basic deep learning methods.
- Fitting and overfitting. Big Data allows fitting very flexible models. This creates the danger of overfitting, which leads to poor out of sample predictability. Controlling overfitting is one of the central tasks in the analysis of Big Data.

4. Literature

The course is built around the textbook "An Introduction to Statistical Learning with Applications in R" (ISL) by James, Witten, Hastie and Tibshirani. The book as well as datasets and practice exercises are available at <https://statlearning.com/>.

5. Programming Language

R/RStudio will be used for the course. While prior knowledge of R is recommended, it is not required. We will provide a detailed introduction to R in the first week of the course (see below).

6. Organization

We will use *OpenOlat* as the organization and communication medium. We advise you to consult *OpenOlat* regularly.

Lectures take place on campus and will contain both, teaching of the theory of machine learning techniques as well as applying the techniques using R. So please, for each session bring your laptop so that you can work on the data problems I provide to you in class. We will also flip the classroom and I will ask you to present your findings of the programming exercises.

Introduction to programming in R: In the second lecture of the course, on April 11, I will provide an introduction to programming in R. This introduction is for students who don't have any prior programming knowledge and we will go through the basics that are required to solve machine learning problems in R. After this session, I will provide coding exercises which you are asked to work on at home.

I strongly encourage all students to work on these exercise right away in the first week to be well prepared for the coming lectures. There will be no lecture on April 18 and I encourage students to use the time of the lecture to work on the programming exercises instead. After this, all students should be well prepared to work on their first machine learning problems in R.

7. Grading

The final grade is based on the take home exam. The exam starts at **Friday, July 28, 2023 at 8.00** and ends at **Friday, July 28, 2023 at 23.59**. So you have 16 hours to work on the exam. In this exam you will be asked to solve data problems using machine learning techniques in R and provide slides which summarize your results. You must hand in two things: a slide deck with the results in pdf format as well as the R codes which generate the results in the pdf. For the grading, we will only consider the results in the slides. However, we will run the codes and if they produce errors or do not produce the same results as in the report, you will get 0 points for the respective exercise.

The style and layout of the slides will also matter for the grade. Presenting and explaining the outcomes of machine learning techniques is one of the major tasks of a data scientist and is an important skill in itself. Hence, we will take this into account for the final grade. For example, please make sure that all graphs and tables are formatted nicely (taking screen-

shots of R outputs is not appropriate. Export all tables and figures to powerpoint/latex and make sure to give clear interpretations. The report should be similar (apart from the appendices) to what you would present in a corporate board meeting, or to your manager when you start working.

8. Office hours

I do not have pre-scheduled office hours and I encourage you to use the ment.io discussions to ask your questions. However, if you have a questions that you cannot post online, please write me an email to [ole.wilms « at » uni-hamburg.de](mailto:ole.wilms@uni-hamburg.de) and we can schedule a meeting:

9. Schedule (tentative)

Interactive lectures including coding exercises:

- April 4: Introduction to machine learning
- April 11: Intro to R
- April 18: **No lecture** (practice R exercises)
- April 25: Introduction to supervised learning and linear regressions
- May 2: Classification problems
- May 9: Resampling methods (cross-validation)
- May 23: Problem Set 1
- May 30: Shrinkage methods (lasso and ridge regressions)
- June 6: Tree-based methods
- June 13: Tree-based methods
- June 20: Problem Set 2
- June 27: Deep learning (neural networks)
- July 4: Deep learning (neural networks)
- July 11: Problem Set 3

Introductory programming session:

- April 11: Introduction to programming in R
- April 18: No lecture, solve R practice exercises

10. Important dates**Take Home Exam:**

- Start date for the take home exam: **Friday, July 28, 2023 at 8.00**
- Deadline for submitting the take home exam: **Friday, July 28, 2023 at 23.59**

Take Home Resit:

- Start date for the take home resit: **Friday, September 22, 2023 at 8.00**
- Deadline for submitting the take home resit: **Friday, September 22, 2023 at 23.59**