

Syllabus

Data Science and Machine Learning 2. (Tools)

- **Instructor:** Zoltan Papp, Jeno Pal (TA)
- **Credits:** 2 (4 ECTS)
- **Term:** Winter 2017-2018
- **Course level:** MSc
- **Prerequisites:** Data Analysis 1,2,3 and Data Science and Machine Learning 1. (Concepts)

Course availability

Student cap: 45. The course is core for MSc in Business Analytics students, they will have priority 1 upon registration. This is also an elective on the MSc in Finance program, so Finance students will be granted priority 2. Students from the waiting list will be let in the class until cap is reached based on their academic/professional background, decided by the instructor.

Course description

This course will build on the previous one (which introduced the basic concepts in machine learning) and will discuss state-of-the-art algorithms for supervised learning (linear models, lasso, decision trees, random forests, gradient boosting machines, neural networks, support vector machine, deep learning etc.). A large part of the course will be dedicated to using (hands-on) the software tools for machine learning used by data scientists in practice (various high-performance R packages, xgboost, libraries for deep learning etc.).

Learning outcomes

Students will get a basic understanding of the main algorithms used in machine learning and will be able to use the main open source libraries implementing these algorithms in practical applications.

Reading list

Available on the course website (for each week), see below.

Assessment

Grading:

- 45% Weekly Assignments (homework exercises). These will be submitted using Moodle.
- 45% Final Exam (closed book)
- 10% Quizzes at the beginning of each lecture, except the first lectures of each course. Missing a lecture or being late will result in 0% of the actual quiz score.

Weekly assignment acceptance policy and achievable grades:

- 100% until the due date
- 50% within 24 hours past the due date

- 0% after that.

Grading policy

Students shall not miss more than 1 day of lectures/seminar (out of 8 days). Failing to do so will yield an administrative fail grade. (If you have a major impediment please contact the Instructor.) To pass, students will need to get at least 60% of the overall grade. Failure to do so, will yield a Fail grade.

Course schedule and materials for each session

Week 9: A deeper understanding of linear models and lasso. Understanding and tuning parameters for trees, random forests and gradient boosting machines. Impact of correlated features. Support vector machines. Tools: R packages, xgboost, lightgbm.

Week 10: Neural networks and deep learning. Reinforcement Learning. Evolutionary Computing. Tools: R packages, Keras.

Week 11: Ensembles, Stacking. Deploying machine learning models to production.

Week 12: Recap and **Final Exam (ML #1 and #2)**.

For any updates see the course website:

<https://github.com/pappzoltan/teach-ML-CEU-master-bizanalytics>