

Syllabus

Data Science and Machine Learning 1. (Concepts)

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

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

After an overview of the entire data science landscape, this course will focus on machine learning. The course will introduce the main fundamental concepts in machine learning (supervised learning, training, scoring, accuracy measures, test set, overfitting, cross-validation, model capacity, hyperparameter tuning, grid and random search, regularization, ensembles, model selection etc.) The concepts will be illustrated with R code, therefore, it requires prior familiarity with R.

Learning outcomes

Students will get a basic understanding of the main concepts in machine learning. They will be prepared for the next course *Data Science and Machine Learning 2 (Tools)*

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 dates and deadlines:

- Jan 16, due on Jan 22 8:00 AM
- Jan 23, due on Jan 29 8:00 AM
- Jan 30, due on Feb 5 8:00 AM

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 and 1 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 1: Overview of data science. The elements of a data science project. Data preparation. Exploratory data analysis, data visualization. Machine learning. Workflow, reproducibility, and productivity.
- Week 2: Introduction to supervised learning. Linear models vs k-nearest neighbors. Training and test error. Validation set. Bias and variance. Model evaluation and selection. Benchmark model selection. Bootstrapping, resampling. Time series split up by time.
- Week 3: A high-level introduction to RF, Lasso, SVM, Boost, NN. Overfitting, regularization, cross-validation, naive implementation. ROC curve, AUC. Calibration plot. Hyperparameter tuning, grid, and random search.
- Week 4: Unsupervised learning. Clustering (k-means, hierarchical). PCA / ICA. Time Series Analysis: correlation, autocorrelation, Filtering. Reviewing best practices in data science and practical research.

For any updates see the course website:

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