

Data Analysis and Machine Learning with Numerical Projects

**Department of Geoscience, Department of Mathematics and
Department of Physics, University of Oslo**

Planned start: Fall semester 2018

Data analysis and machine learning: a new 10-ECTS course for both CS students and others

A much needed course across disciplines.

Pre-requirement: basic knowledge in programming and numerics. Required courses are the equivalents to the University of Oslo mathematics courses MAT1100, MAT1110, MAT1120 and at least one of the corresponding computing and programming courses INF1000/INF1110 or MAT-INF1100/MAT-INF1100L/BIOS1100/KJM-INF1xxx.

Learning outcomes

- Learn about basis data analysis, Bayesian statistics, Monte Carlo methods, data optimization and machine learning
- Capable of extending the acquired knowledge to other systems and cases
- Have an understanding of central algorithms used in data analysis and machine learning
- Basic knowledge of Bayesian statistics and learning and common distributions
- Knowledge of central aspects of Monte Carlo methods, Markov chains, Gibbs samplers and their possible applications, from numerical integration to simulation of stock markets.
- Understanding of linear methods for regression and classification
- Neural network, genetic algorithms and Boltzmann machines
- Numerical projects to illustrate the theory play a central role and students are expected to know modern programming languages like Python or C++.

The course has two central parts

The course will have two main parts:

1. Statistical analysis and optimization of data
2. Machine learning

Computational aspects play a central role and the students are expected to work on numerical projects which illustrate the theory. Some of the projects can be coordinated with the high-performance programming course (course code to be added).

Statistical analysis and optimization of data. The following topics will be covered

1. Basic concepts, expectation values, variance, covariance, correlation functions and errors
2. Simpler models, binomial distribution, the Poisson distribution, simple and multivariate normal distributions
3. Central elements of Bayesian statistics and modeling
4. Monte carlo methods, Markov chains, Metropolis-Hastings algorithm, ergodicity
5. Linear methods for regression and classification
6. Estimation of errors using blocking, bootstrapping and jackknife methods
7. Practical optimization using Singular-value decomposition and least squares for parameterizing data

All topics will be supported by examples, hands-on exercises and project work.

Machine learning. The following topics will be covered

1. Gaussian and Dirichlet processes
2. Boltzmann machines
3. Neural networks
4. Genetic algorithms

All topics will be supported by examples, hands-on exercises and project work.

Practicalities

1. Four lectures per week, Fall semester, 10 ECTS
2. Four hours of laboratory sessions for work on computational projects
3. Three projects which are graded and count 60% of the final grade
4. A selected number of weekly assignments which count 10% of the final grade
5. Final written exam which counts 30% of the final grade
6. Organized by the Departments of Geoscience, Mathematics and Physics
7. Possible teachers first time: John Burkhart, Morten Hjorth-Jensen and XXX
8. The course is part of the CS Master of Science program, but is open to other bachelor and Master of Science students at the University of Oslo.
9. Grading scale: Grades are awarded on a scale from A to F, where A is the best grade and F is a fail.
10. The course will be offered as a 4XXX and 3XXX course.

Possible textbooks

General learning book on statistical analysis:

1. Christian Robert and George Casella, Monte Carlo Statistical Methods, Springer
2. Peter Hoff, A first course in Bayesian statistical models, Springer

General Machine Learning Books:

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press
2. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer
3. David J.C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press
4. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning, Springer
5. David Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press