Week 34: Introduction to the course, Logistics and Practicalities

Morten Hjorth-Jensen[1,2]

Department of Physics and Center for Computing in Science Education,
University of Oslo, Norway[1]

Department of Physics and Astronomy and Facility for Rare Isotope Beams,
Michigan State University, USA[2]

Week 34, August 21-25, 2023

Overview of first week

1. The sessions on Tuesdays and Wednesdays last four hours for each group (four groups in total) and will include lectures in a flipped mode (promoting active learning) and work on exercices and projects.

2. The sessions will begin with lectures, discussions, questions and answers about the material to be covered every week. Videos and teaching material will be announced in due time.

3. There are four groups:
   - Tuesdays 815am-12pm and 1215pm-4pm
   - Wednesdays 815am-12pm and 1215pm-4pm.

4. On Thursdays we have a regular lecture. These lectures start at 1215pm and end at 2pm and serve the aims of giving an overview over various topics. These lectures will also be recorded.

The labs are also available till 6pm Tuesdays and Wednesdays. Videos and learning material with reading suggestions will be made available before each week starts.

Schedule first week

- ▶ August 22: Presentation of the course, aims and content. Introduction to software and repetition of Python Programming, linear algebra and basic elements of statistics. Please select group.
- ▶ August 23: Presentation of the course, aims and content. Introduction to software and repetition of Python Programming, linear algebra and basic elements of statistics. Please select group.
- ▶ August 24: Lecture: Linear regression, examples and theory

Lectures and ComputerLab

- The sessions on Tuesdays and Wednesdays last four hours and will include partly lectures in a flipped mode (promoting active learning) and work on exercices and projects.
- Thursdays: regular lectures (12.15pm-2pm)
- Weekly reading assignments and videos needed to solve projects and exercises.
- Weekly exercises. You can hand in exercises if you want and get an extra score, see below.
- Detailed lecture notes, exercises, all programs presented, projects etc can be found at the homepage of the course.
- Weekly plans and all other information are on the official website. This info will also be conveyed via weekly emails.
- No final exam, three projects that are graded and have to be approved.

Communication channels

- Chat and communications via `canvas.uio.no`
- **Discord** channel will be added asap

Course Format

- ▶ Three compulsory projects. Electronic reports only using Canvas to hand in projects and git as version control software and GitHub for repository (or GitLab) of all your material.
- ▶ Evaluation and grading: The three projects are graded and each counts 1/3 of the final mark. No final written or oral exam.
  1. For the last project each group/participant submits a proposal or works with suggested (by us) proposals for the project.
  2. If possible, we would like to organize the last project as a workshop where each group presents this to all other participants of the course
  3. Based on feedback etc, each group finalizes the report and submits for grading.
- ▶ Python is the default programming language, but feel free to use C/C++, Julia and/or Fortran or other programming languages. All source codes discussed during the lectures can be found at the webpage and github address of the course.

Teachers

- Morten Hjorth-Jensen, morten.hjorth-jensen@fys.uio.no
  - **Phone**: +47-48257387
  - **Office**: Department of Physics, University of Oslo, Eastern wing, room FØ470
  - **Office hours**: *Anytime*! Individual or group office hours can be arranged either in person or via zoom. Feel free to send an email for planning.

- Ida Torkjellsdatter Storehaug, i.t.storehaug@fys.uio.no

- Fahimeh Najafi, fahimeh.najafi@fys.uio.no

- Mia-Katrin Ose Kvalsund, m.k.o.kvalsund@fys.uio.no

- Karl Henrik Fredly, k.h.fredly@fys.uio.no

- Adam Jakobsen, adam.jakobsen@fys.uio.no

- Daniel Haas Beccatini Lima, d.h.b.lima@fys.uio.no

Deadlines for projects (tentative)

1. Project 1: October 9 (available September 4) graded with feedback)
2. Project 2: November 6 (available October 6, graded with feedback)
3. Project 3: December 11 (available November 10, graded with feedback)

Extra Credit (not mandatory), weekly exercise assignments, 10 in total (due Friday same week), 10% additional score. The extra credit assignments are due each Friday and can be uploaed to **Canvas** in your preferred format (although we prefer jupyter-notebooks). First assignment is for week 35. Each weekly exercise set counts 1%.

Grading

Grades are awarded on a scale from A to F, where A is the best grade and F is a fail. There are three projects which are graded and each project counts 1/3 of the final grade. The total score is thus the average from all three projects.

The final number of points is based on the average of all projects and the grade follows the following table:

- ▶ 92-100 points: A
- ▶ 77-91 points: B
- ▶ 58-76 points: C
- ▶ 46-57 points: D
- ▶ 40-45 points: E
- ▶ 0-39 points: F-failed

In addition you can get an extra 10% score for weekly assignments (10 in total and due each Friday). Each weekly assignment counts 1%.

Reading material

The lecture notes are collected as a jupyter-book at
https://compphysics.github.io/MachineLearning/doc/
LectureNotes/_build/html/intro.html.
In addition to the lecture notes, we recommend the books of
Bishop, Hastie et al, Murphy and Goodfellow et al. We will follow
these texts closely and the weekly reading assignments refer to
these texts. The text by Hastie et al is also widely used in the
Machine Learning community. Finally, we also recommend the
hands-on text by Geron, see next slide for links.

Textbooks

- Goodfellow, Bengio, and Courville (GBC), Deep Learning
- Christopher M. Bishop (CB), Pattern Recognition and Machine Learning
- Hastie, Tibshirani, and Friedman (HTF), The Elements of Statistical Learning, Springer.
- Aurelien Geron (AG), Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly. This text is very useful since it contains many code examples and hands-on applications of all algorithms discussed in this course.
- Kevin Murphy (KM), Probabilistic Machine Learning, an Introduction

Reading suggestions week 34

This week: Refresh linear algebra, GBC chapters 1 and 2. HTF chapters 2 and 3. Install scikit-learn. See lecture notes for week 34 at `https://compphysics.github.io/MachineLearning/doc/web/course.html` (these notes).

Prerequisites

Basic knowledge in programming and mathematics, with an emphasis on linear algebra. Knowledge of Python or/and C++ as programming languages is strongly recommended and experience with Jupiter notebook is recommended. 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-INF1100. Most universities offer nowadays a basic programming course (often compulsory) where Python is the recurring programming language.

Topics covered in this course: Statistical analysis and optimization of data

The course has two central parts

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

These topics will be scattered thorughout the course and may not necessarily be taught separately. Rather, we will often take an approach (during the lectures and project/exercise sessions) where say elements from statistical data analysis are mixed with specific Machine Learning algorithms.

Statistical analysis and optimization of data

We plan to cover the following topics:

▶ Basic concepts, expectation values, variance, covariance, correlation functions and errors;

▶ Simpler models, binomial distribution, the Poisson distribution, simple and multivariate normal distributions;

▶ Central elements of Bayesian statistics and modeling;

▶ Gradient methods for data optimization;

▶ Monte Carlo methods, Markov chains, Gibbs sampling and Metropolis-Hastings sampling (tentative);

▶ Estimation of errors and resampling techniques such as the cross-validation, blocking, bootstrapping and jackknife methods;

▶ Principal Component Analysis (PCA) and its mathematical foundation;

Machine Learning

- ▶ Pre deep-learning revolution (2008 approx)
    - ▶ Linear Regression and Logistic Regression, classification and regression problems;
    - ▶ Bayesian linear and logistic regression, kernel regression;
    - ▶ Decisions trees, Random Forests, Bagging and Boosting methods;
    - ▶ Support vector machines (only survey);
    - ▶ Unsupervised learning and dimensionality reduction, from PCA to clustering;

- ▶ Deep learning
    - ▶ Neural networks and deep learning;
    - ▶ Convolutional neural networks;
    - ▶ Recurrent neural networks;
    - ▶ Autoencoders
    - ▶ Generative methods with an emphasis on Boltzmann Machines, Variational Autoencoders and Generalized Adversarial Networks;

Hands-on demonstrations, exercises and projects aim at deepening your understanding of these topics.

Extremely useful tools, strongly recommended

and discussed at the lab sessions

▶ GIT for version control, and GitHub or GitLab as repositories, highly recommended. This will be discussed during the first exercise session

▶ Anaconda and other Python environments, see intro slides and links to programming resources at
https://computationalscienceuio.github.io/
RefreshProgrammingSkills/intro.html

Other courses on Data science and Machine Learning at UiO

- ▶ FYS5419 Quantum Computing and Quantum Machine Learning
- ▶ FYS5429 Advanced Machine Learning for the Physical Sciences
- ▶ STK2100 Machine learning and statistical methods for prediction and classification.
- ▶ IN3050/4050 Introduction to Artificial Intelligence and Machine Learning. Introductory course in machine learning and AI with an algorithmic approach.
- ▶ STK-INF3000/4000 Selected Topics in Data Science. The course provides insight into selected contemporary relevant topics within Data Science.
- ▶ IN4080 Natural Language Processing. Probabilistic and machine learning techniques applied to natural language processing.

Other courses on Data science and Machine Learning at UiO, contn

- ► STK-IN4300 Statistical learning methods in Data Science. An advanced introduction to statistical and machine learning. For students with a good mathematics and statistics background.
- ► IN3310/4310 Deep Learnig for Image Analysis
- ► STK4051 Computational Statistics
- ► STK4021 Applied Bayesian Analysis and Numerical Methods

Learning outcomes

This course aims at giving you insights and knowledge about many of the central algorithms used in Data Analysis and Machine Learning. The course is project based and through various numerical projects, normally three, you will be exposed to fundamental research problems in these fields, with the aim to reproduce state of the art scientific results. Both supervised and unsupervised methods will be covered. The emphasis is on a frequentist approach, although we will try to link it with a Bayesian approach as well. You will learn to develop and structure large codes for studying different cases where Machine Learning is applied to, get acquainted with computing facilities and learn to handle large scientific projects. A good scientific and ethical conduct is emphasized throughout the course.

Learning outcomes, continued

More specifically, after this course you will

- Learn about basic data analysis, statistical analysis, Bayesian statistics, Monte Carlo sampling, data optimization and machine learning;
- Be capable of extending the acquired knowledge to other systems and cases;
- Have an understanding of central algorithms used in data analysis and machine learning;
- Understand linear methods for regression and classification, from ordinary least squares, via Lasso and Ridge to Logistic regression;
- Learn about neural networks and deep learning methods for supervised and unsupervised learning. Emphasis on feed forward neural networks, convolutional and recurrent neural networks;
- Learn about about decision trees, random forests, bagging and