Projects
Machine Learning Course
Fall 2016

EPFL
School of Computer and Communication Sciences
Martin Jaggi & Rüdiger Urbanke
mlo.epfl.ch/page-136795.html
epfmlcourse@gmail.com

Course Overview - CS-433

Course Goals

Define the following basic ML problems and explain the main differences between them:

Regression, classification, clustering, dimensionality reduction, neural networks, ...

Describe a few important models and algorithms for the basic ML problems.

Implement and apply these methods to real-world problems.

Compare the methods and choose one of them.

Criticize and defend your choice of method.

Understand and apply the theory behind ML methods taught in the course and generalize them to new problems.

Continue to work through difficulties or initial failure to find optimal solutions.

Assess one's own level of skill acquisition, and plan on-going learning goals.

Syllabus

We will cover the following ML methods and concepts

- Linear Regression
- Optimization
- Cross-validation, overfitting, bias-variance trade-off
- Classification
- Clustering
- Matrix Factorizations
- Neural Networks
- Random Forests, Bayesian methods

The syllabus provided on the website is more precise but subject to change, especially for the later parts of the lecture.

What not to expect

You will not learn ALL advanced methods.

You will not learn ALL the details.

This course is not about big data or large-scale methods.

This is not a course about numerical optimization, neither is it about statistics. We will use both of these and learn basic techniques only.

We will not teach the pre-requisite for ML. You have to learn that on your own.

This course does not teach you all that you need to know to be able to apply machine learning, but this course will get you started for sure.

Labs

Weekly every Thursday 14:15 - 16:00, in the following rooms:
INF119 (A-Ch); INJ218 (Ci-Kj); INM11 (Ko-Pe); INM202 (Pi-Z), according to last-name.

All labs and projects will be in *Python* this year. See the first lab to get started.

Prerequisites

We will revise some of the concepts in the first and second weeks of the course. However, you should still read these on your own so that you can follow the course well. Here is a list of pre-requisites.

Vector and Matrix Algebra. Vector and matrix multiplication, matrix inversion and determinants, rank, null and range space, eigenvalue decomposition. Refer to first year courses, or Gilbert Strang's book for example.

Vector and Matrix Calculus. Important: The definition of derivative with respect to vectors and matrices. For reference, see the Matrix Cookbook.

Scientific Computing Languages. Python Basics (see tutorial in the first lab).

Probability and Statistics. Conditional and joint distribution, independence, Bayes' rule, random variable and expectation, law of large numbers. Read Chris Bishop's book (Chapter 2).

Gaussian Distribution. Univariate and multivariate, conditional, joint and marginals. See again Chris Bishop's book (Chapter 2).

Writing Scientific Documents using Latex (not necessary but preferred). Many tutorials are available online.

Resources

Course Webpage

We will use the course website for all material, except code

mlo.epfl.ch/page-136795.html

Code templates and solutions for the labs will be made available on our github repository

github.com/epfml/ML_course

Active Participation / Clicker / Discussions

For some active participation in the lectures, please point your browser to this speak-up room (clicker).

We provide PDF lecture notes on the website and additionally also on Nota Bene so you can comment & discuss them.

Lecture Notes

During lectures, we will mainly use Emtiyaz Khan's teaching-award-winning lecture material.

His notes provide blank writing space on the right hand side. The PDFs for each lecture will be available on the website before the day of the lecture, and will often be annotated during the lecture.

Recommended Textbooks

The following books contain relevant material to the course:

- S. Shalev-Shwartz and S. Ben-David: Understanding Machine Learning - From Theory to Algorithms.
- G. James, D. Witten, T. Hastie and R. Tibshirani: An introduction to statistical learning.
- T. Hastie, R. Tibshirani and J. Friedman: Elements of statistical learning.
- C. Bishop: Pattern Recognition and Machine Learning.

- K. Murphy: Machine Learning: A Probabilistic Perspective.

The first three books are available online for free.

You do not have to buy any books, since we will only refer to few chapters in these. Some physical copies are available in the library.

Assessment and Practical Projects

- Project 1 (10%), due Oct 31st
- Project 2 (30%), due Dec 22nd
- Final exam (60%)

Project 1 (10%)

The goal of this project is to help you prepare for Project 2.

In project 1, you will work in a small group of 3 students (2 in exceptional cases).

You will implement the most important methods covered in the lectures and labs so far.

Additionally, we will provide you with an interesting real-world dataset, and organize our own competition here:

inclass.kaggle.com/c/epfml-project-1

A detailed project description will be posted on the website very soon.

For this small first project, the groups will be assigned randomly by us. The assignment for project 2 might be different. Either chosen by you, or random, this is not decided yet.

You will also submit your Python code, and a 2 page PDF report. *Deadline: Oct 31st.*

Project 2 (30%)

Project 2 is the final project and gives you more freedom and responsibilities.

Again, you will work in a group of 3 people.

You can freely choose between three challenges with real-word data problems:

inclass.kaggle.com/c/epfml-rec-sys
inclass.kaggle.com/c/epfml-text
inclass.kaggle.com/c/epfml-segmentation

Submitting your predictions to the kaggle platform allows you to get immediate feedback on your performance.

Finally, you will also submit your Python code, and a 4 page PDF report. *Deadline: Dec 22nd.*

Final exam (60%)

A very standard final exam.

It will contain questions on what you have learned during the lectures and exercise sessions.

We will give you a sample exam before for you to practice.

You are allowed to bring one cheat sheet (A4 size paper, both sides can be used), either handwritten or 11 point minimum font size.

No calculator, No collaborations. No cell phones. No laptops etc.

Contact Us

Email to epfmlcourse@gmail.com to contact the teachers and TAs.

Teaching Assistants:

Mohamad Dia

Ksenia Konyushkova

Victor Kristof

Taylor Newton

Farnood Salehi

Benoît Seguin

Frederik Kunstner (AE)

Fayez Lahoud (AE)

Tao Lin (AE)

Arnaud Miribel (AE)

Vidit Vidit (AE)

The TAs will be helping you during the exercise sessions and projects.

Credits

Teaching material mostly by Emtiyaz Khan.
Python code translation by Tao Lin and other TAs. Additional material by Carlos Becker, Matthias Seeger, Yannic Kilcher, Aurelien Lucchi, Rüdiger Urbanke, Martin Jaggi and many others...