Statistical Machine Learning

Course Information and Foreword

MATH-412 - Statistical Machine Learning

Course key facts

- Lectures : Wed 8 :15-10 in MA A331
- Lecturers and TA :
  - Guillaume Obozinski (Swiss Data Science Center)
  - YunHo (PhD student, Statistics department)
  - Hannes Gubler (master student, Math department)
- Exercises sessions : Th 10-12 in MA A112
  - Number of key examples and important ideas seen as exercises
  - Working on the exercises is key to succeed at the exam
  - Some in exercises in R/Python (or similar). Solutions provided in R.
- Evaluation :
  - Final exam 70% (questions similar to exercise sessions)
  - Project 30% (article to read on classical technique or algorithm+numerical investigation)
- Moodle
  - Slides + link to the video channel
  - Exercise sheets + Solutions (later)

Evaluation

Final Exam (70%) :

- 3 hours written in class in January
- short questions + problems

Projects (30%) :

- Groups of 3
- To be chosen mid-semester
- Research or review article to read on classical technique/algorithm
- + some implementation / application on data (Python/R or other)
- Goal : show that you can leverage your understanding of the concepts of the course to understand and use well new methods.
- Very short presentation on Dec 6 and 7th
- 4 pages report due in January

Syllabus (Tentative)

Sep 20 1. Intro, Decision theory, Linear regression
Sep 27 2. Linear model
Oct 04 3. Regularization
Oct 11 4. Local averaging, binary classification + List of projects available
Oct 18 5. Linear binary classification,
Oct 25 6. Support Vector Machines
Nov 01 7. Kernelized methods Team formed + Projects selected
Nov 08 8. Decision trees,
Nov 15 9. Random Forests + Help session on project
Nov 22 10. Boosting
Nov 29 11. PCA
Dec 06 12. Project presentations
Dec 13 13. Clustering, EM
Dec 20 14. Deep learning

Style of the course

- Laying foundations of Machine Learning
- Mathematical approach
- Statistical perspective
- Goal : understanding the underlying principles behind the algorithm
- Not a course that teaches you how to use libraries

Organisation

- The following info is further detailed on the Moodle page :
- **Videos on EPFL mediaspace** (almost the) Entire course in video
- **Warning :** some of the lectures available as videos will not be covered this year in class. The corresponding material is completely optional.
  - $\rightarrow$ Model selection (was lecture 3) will be not covered this year.
- **Ed** We will communicate via Ed on Moodle.

Intro...

What kind of learning ?

Learn to :

- Recognize different kinds of butterflies from specimens
- Detect pedestrians on the street with an on board camera
- Read postal codes/checks
- Produce the syntactical relations between words in a sentence
- Predict which chemical components can react with a given protein
- Translate from a language to another
- Recognize speech
- Fly a helicopter

Learn *empirically* from a flow of experience, i.e. from a data stream

Early machine learning

Arthur Lee Samuel testing his checkers-playing program on the IBM 704

He coined the expression "machine learning" in 1959.

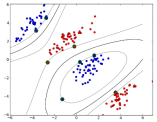
A definition of machine learning

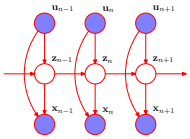
"A computer program is said
to
learn from experience $\mathcal{E}$
with respect to some class of tasks $\mathcal{T}$
and performance measure $\mathcal{P}$
if
its performance at tasks in $\mathcal{T}$,
as measured by $\mathcal{P}$,
improves with experience $\mathcal{E}$*"*.

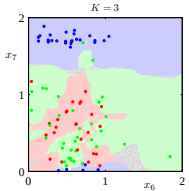
Tom Mitchell
chair of CMU ML dpt.

A broad spectrum of techniques

- Empirical risk minimization
- Local averaging
- Kernel methods
- Probabilistic modeling and Bayesian methods
- Neural networks & deep learning
- Decision trees
- Ensemble methods
- PAC Bayesian techniques







Syllabus (Tentative)

Sep 20 1. Intro, Decision theory, Linear regression
Sep 27 2. Linear model
Oct 04 3. Regularization
Oct 11 4. Local averaging, binary classification + List of projects available
Oct 18 5. Linear binary classification,
Oct 25 6. Support Vector Machines
Nov 01 7. Kernelized methods Team formed + Projects selected
Nov 08 8. Decision trees,
Nov 15 9. Random Forests + Help session on project
Nov 22 10. Boosting
Nov 29 11. PCA
Dec 06 12. Project presentations
Dec 13 13. Clustering, EM
Dec 20 14. Deep learning