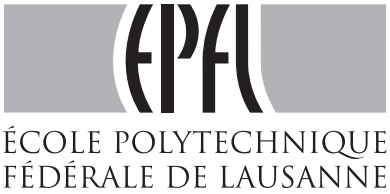
Course Information

Mohammad Emtiyaz Khan

September 15, 2014



Contents

1 Lectures and Labs

For lectures, bring printed notes and your smartphone (or a laptop) for in-class quizzes.

- Lecture (8h15-10h) on Tuesdays in CE-1
- Lecture (8h15-10h) on Thursdays in CE-4.

For labs, bring your laptops.

- Lab (14h15-16h) on Thursday in INJ-218 (when full, go to INM 202).

2 Course Summary

2.1 Course Goals

- Define the following basic ML problems and explain the main differences between them:
  - Regression, classification, clustering, dimensionality reduction, time-series.
- 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.
- Critique and defend your choice of method.
- Derive 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 their on-going learning goals.

2.2 Syllabus

We will cover the following 4 types of ML methods.

- [12 lectures] Regression and classification.
- [4 lectures] Clustering.
- [4 lectures] Dimensionality reduction.
- [4 lectures] Time-series analysis.

2.3 A rough outline of the course

- [4 lectures] Regression using Linear regression.
  - Revise basics and learn to implement and test ML methods.
- [2 lectures] Cross-validation, overfitting, bias-variance trade-off.
  - The most important fundamentals of ML.
- [2 lectures] Classification using Logistic regression.
  - Generalize the theory to a new problem.
  - This is an important exercise for exam.

- [6 lectures] Advanced Classification and Regression methods. We will pick only some (NOT all) of the following:
  - Generalized linear model and Multi-class Classification.
  - Curse of dimensionality and k-NN.
  - Kernel methods and SVM.
  - Tree based methods and Decision Trees.
  - Bayes rule and Naive Bayes Classifier.
  - Bayesian methods and Gaussian Processes.
  - Deep learning and Neural Networks.
  - Bayes decision theory and LDA.
- [4 lectures] Clustering.
  - Mixture Models and k-means clustering.
  - Gaussian Mixture Model and Expectation Maximization (EM) algorithm.
- [4 lectures] Dimensionality reduction
  - PCA and ICA
- [4 lectures] Time-series analysis
  - Kalman filters and Hidden Markov Model (HMM).

Please remember that this is a tentative schedule. We will NOT cover everything in the classification and regression section. Since we have only 6 lectures for advanced methods, we will only be able to cover a subset of those methods.

2.4 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.

2.5 Pre-requisite

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.

- Matrix algebra.
  - Vector and matrix multiplication, matrix inversion and determinants, rank, null and range space, eigenvalue decomposition.
  - Refer to Gilbert Strang's book. Also, read the handout posted in the Moodle about matrix algebra.
- Matrix calculus.
  - Definition of derivative with respect to vectors and matrices.
  - We will use Matrix Cookbook for formula. Download at http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf.
- Scientific computing languages.
  - Matlab Basics.

  - Read Kevin Murphy's Matlab tutorial, available at http://ubcmatlabguide.github.io/.

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

- Gaussian distribution.
  - Univariate and multivariate, conditional, joint and marginals.
  - Read Chris Bishop's book (Chapter 2).

- Writing scientific documents using Latex (not necessary but preferred).
  - Many tutorials available online.

3 Resources

3.1 Course Webpage and Moodle

We will use the course website for most things: http://icapeople.epfl.ch/mekhan/pcml15.html.

We will use Moodle for group assignments, news forum, posting data etc.

3.2 Lecture notes

During lectures, I will use lecture notes where you have to fill in the blanks. These will be available on Moodle before the lecture. Please print it and bring it to the class. An annotated copy will be available after the lecture. Sometimes, I will use slides too. Again, annotated slides will be on Moodle.

3.3 Books

I will use the following books to teach. You will be required to read few chapters of these books (I will let you know about this too).

- G. James, D. Witten, T. Hastie and R. Tibshirani: An introduction to statistical learning (free download from http://www-bcf.usc.edu/~gareth/ISL/).

- T. Hastie, R. Tibshirani and J. Friedman: Elements of statistical learning (download from http://statweb.stanford.edu/~tibs/ElemStatLearn/).
- C. Bishop: Pattern Recognition and Machine Learning.
- K. Murphy: Machine Learning: A Probabilistic Perspective.

The first two books are available online for free. You do not have to buy these books, since we will only refer to few chapters in these. There are some copies in the library.

3.4 In-Class Quizzes

We will use ResponseWare tool for in-class quizzes using which you can take quizzes online at http://www.rwpoll.com.

Participating in in-class quiz: In each class, I will give a session number required to access the quiz at the website http://rwpoll.com. When I ask a question, it will automatically appear on your screen and you will be able to choose your answer.

How can you participate? There are several ways. Easiest way is to use a smartphone or any

other networked device like a laptop. You can submit your answers via a web browser. Remember that you need an internet connection for this to work.

You can also use the corresponding App (only) on iPhone and Android. Remember to set your region to 'North America', otherwise the app doesn't work.

More details at http://clickers.epfl.ch/cms/site/clickers/lang/en/using-your-smartphone.

4 Evaluations

- (10%) Project 1.
- (30%) Project 2.
- (60%) Final exam.

4.1 Project 1 (worth 10%)

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

- Work in groups of at most 2 people.
- The group assignment cannot be changed and will remain the same for Project 2.
- We will provide you two datasets and a list of algorithms to use.
- The data will be synthetic where we (I and the TAs) know the ground truth.
- Each group might get their own data from a different ground truth, so you can not cheat.
- You have about 3 weeks to turn in your results.

See the project page from last year to get an idea:
http://cvlabwww.epfl.ch/~cjbecker/pcml_project1/project1-summary.html

4.2 Project 2 (worth 30%)

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

- Work in groups of at most 2 people.
- We will give you a list of projects (with data and task).
- Again each group might get their own data.
- You have about 5 weeks to turn in your results.

Here is a schedule from LAST YEAR, and we will follow very similar schedule this year as well.

Tue Sep 30	20:00	Project 1 details posted
Tue Oct 7	20:00	Project 1 data posted
Mon Oct 13	14:00	Project 1 groups details due
Mon Oct 27	14:00	Project 1 due
Tue Nov 4	20:00	Project 2 details posted
Tue Nov 11	20:00	Project 2 data posted
Mon Nov 17	14:00	Project 2 group details and project name due
Mon Dec 15	14:00	Project 2 due

4.3 Final exam (worth 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) and a calculator.
- No collaborations. No cell phones. No laptops etc.

5 Teaching Assistants

Email at `pcml.epfl@gmail.com` to contact the teacher and TAs. Do not send personal emails.

- Carlos Becker
- Becker Carlos Joaquin
- Bermdez Chacn Rger
- Newton Taylor Howard
- Salehi Farnood
- Seguin Benot Laurent Auguste
- Victor Kristof
- Dennis Meier (Master)
- Merlin Nimier-David (Master)

TAs will be helping you during your exercise sessions and projects.

Notes