COMP 598 – Applied Machine Learning
Lecture 1: Introduction

Instructor: Joelle Pineau (jpineau@cs.mcgill.ca)

TAs: Pierre-Luc Bacon (pbacon@cs.mcgill.ca)
     Ryan Lowe (ryan.lowe@mail.mcgill.ca)

Class web page: www.cs.mcgill.ca/~jpineau/comp598

Outline for today

• Overview of the syllabus
• What will this course contain?
• What is ML?
• Examples of ML applications
Course objectives

• To develop an understanding of the fundamental concepts of ML.

• To emphasize good methods and practices for effective deployment of real systems.

• To acquire hands-on experience with basic tools and algorithms.

About You (as of Sep. 3 2015)

74 registered, primarily from:
- Computer Science
- Mathematics / Statistics / Biostatistics / Epidemiology
- Biomedical, Electrical, Software Engineering

… and a few from:
- Biology, Physics, Cognitive science, Linguistics, Economics,

10% PhD, 40% MSc, 50% B./BSc. candidates.

How many more could not register?
About Joelle Pineau

• What have I done?
  – B.A.Sc. in Engineering (U.Waterloo) 1993 - 1998
  – Assistant / Associate Prof at McGill 2004 - …

• Co-director of the Reasoning and Learning Lab.

• What kind of research do I do?
  – Machine learning (reinforcement learning, deep learning, online learning, representation learning, …)
  – Planning and decision-making
  – Robotics
  – Personalized medicine and health-care

Research areas in my lab
2015


COMP-598: Applied Machine Learning 7 Joelle Pineau
About the course: Tentative list of topics

- Linear regression.
- Linear classification.
- Performance evaluation, overfitting, cross-validation, bias-variance analysis, error estimation.
- Dataset analysis.
- Naive Bayes.
- Decision and regression trees.
- Support vector machines.
- Neural networks.
- Deep learning.
- Unsupervised learning and clustering.
- Feature selection.
- Dimensionality reduction.
- Regularization.
- Data structures and Map-Reduce.
- Ensemble methods.
- Cost-sensitive learning.
- Online / streaming data.
- Time-series analysis.
- Semi-supervised learning.
- Recommendation systems.
- Ranking and preference learning.
- Applications.

About the course

- **During class:**
  - Primarily lectures, with some seminars, paper discussions, problem-solving.

- **Outside of class:**
  - Read papers, watch videos, solve problems, complete four mini-projects, peer review work of colleagues.

IMPORTANT!
These target different, but complementary, knowledge & skills!
About the course

• Prerequisites:
  – Basic knowledge of a programming language (Matlab, R are ok.)
  – Basic knowledge of probabilities/statistics (e.g. MATH-323, ECSE-305).
  – Basic knowledge of calculus and linear algebra.
  – Some AI background is recommended (COMP-424 or ECSE-526) but not required.

• Anterequisites:
  – If you took COMP-652 before 2014, you CANNOT take COMP-598.
  – However taking COMP-652 during/after Winter 2014 is ok (course was re-designed to avoid overlap).

Lecture notes will be available on class website.

Discussions will not!

Evaluation:
• Weekly quizzes and exercises (5%)
• One in-class midterm (35%)
• Four data analysis case studies + peer reviews (60%)

Coursework policy:
All course work should be submitted online (details to be given in class), by 11:59pm, on the assigned due date. Late work will be subject to a 30% penalty, and can be submitted up to 1 week after the deadline.
No make-up quizzes or midterm will be given.
About the course

• Four mini-projects:
  1. Data acquisition and pre-processing.
  2. Classification task #1. (Text)
  3. Classification task #2. (Complex signal, e.g. speech, EEG, image)
  4. Final project. (Variety of datasets)

• Format:
  – Projects will be submitted as written report + working code/data.

• Work to be done in teams of 3. Work with different people each time.

Expectations

• Come to class prepared.
• Participate in discussions and sessions.

• Work hard on the data case studies.
  – Show initiative, creativity, sound analytical skills, good communication skills.
  – Work well in a team.

• Provide constructive feedback in peer-reviews.
• Respect the coursework policy.
Course material

- No mandatory textbook, but a few good textbooks are recommended on the syllabus (some freely available online).

  Shalev-Schwartz & Ben-David

  Hastie, Tibshirani & Friedman.

  Bishop.

Tools

- Many software packages are available, including reasonably general libraries in Java, C++, Python and others.

- Many advanced packages for specialized algorithms.

- Strong push in the community to make software available publically.
Read this carefully

- **Some of the course work will be individual, other components can be completed in groups.** It is the responsibility of each student to understand the policy for each work, and ask questions of the instructor if this is not clear.

- It is the responsibility of each student to **carefully acknowledge all sources** (papers, code, books, websites, individual communications) using appropriate referencing style when submitting work.

- **We will use automated systems to detect possible cases of text or software plagiarism.** Cases that warrant further investigation will be referred to the university disciplinary officers. Students who have concerns about how to properly use and acknowledge third-party software should consult the course instructor or TAs.

**What is machine learning?**

- A definition (by Tom Mitchell):
  
  "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"

- More technically:

  "A computer program is said to learn from experience $E$ with respect to some class of tasks $T$ and performance measure $P$, if its performance at tasks in $T$, as measured by $P$, improves with experience $E"
Case study #1: Optimal character recognition

- Handwritten digit recognition: >99% accuracy (on a large dataset).

![Image of a neural network](image1)

Previously seen known examples

New example to classify

Boxes represent the weights into a hidden node in a neural network learner.

Case study #2: Computer Vision

- Face recognition. Not always perfect!

![Image of faces](image2)

![Image of cookies](image3)
Case study #3: Text analysis

• Learning a language model:

Text corpus \(\rightarrow\) Statistical language model

\[
P(W_n | W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})}
\]

Speech recognition pipeline
Case study #3: Text analysis

- Learning a language model:
  \[ \text{Source Language Text} \xrightarrow{T} \text{model} \]
  \[ \frac{\text{Transformation}}{r_t^i} \]
  \[ \text{Global Search:} \]
  \[ \maximize_{r_t^i} \Pr(r_t^i) \cdot \Pr(t^i | r_t^i) \]
  \[ \text{over } e_t^i \]
  \[ \text{Transformation} \]
  \[ \text{Target Language Test} \]

Machine translation pipeline

Case study #4: The Netflix Prize

**Task:** Improve Netflix's recommendation system by 10%.

**Training data:** 10^8 movie ratings, to build the ML algorithm.

**Test set:** 1.5x10^8 ratings to evaluate final performance.

**Quiz set:** 1.5x10^6 ratings to calculate leaderboard scores.
Types of machine learning

- Supervised learning
  - Classification
  - Regression
- Unsupervised learning
- Reinforcement learning
Final comments

• Come to class! Come prepared!

• For Friday:
  – (Must) Read this paper:
  – (If necessary) Review basic algebra, probability, statistics.
    • Ch.1-2 of Bishop.
    • Many online resources.
  – (Optional) Read Chap.1-2 of Bishop, Ch. 1 of Hastie et al. or Ch.2 of Shalev-Schwartz et al.