CPSC 540 - Machine Learning

Introduction

Mark Schmidt

University of British Columbia

Fall 2014

Location/Dates

- Course homepage:
  `http://www.cs.ubc.ca/~schmidtm/Courses/540`
- Office hours: Tuesday 300-4 (ICCS 193), or by appointment.
- Tutorials: Thursdays 300-4 (FORW 519).
- TA: Mohamed Ahmed.

Motivation

- Machine learning is one the fastest growing areas of science.
- Key idea: use data to solve hard pattern recognition problems.
- Recent successes: Kinect, book/movie recommendation, spam detection, credit card fraud detection, face recognition, speech recognition, object recognition, self-driving cars.
- Many more applications to be discovered!

Prerequisites

There will be some review, but you should know:

- Multivariate calculus:

$$\nabla_x x^T a = a.$$

- Linear algebra:

$$Ax = \lambda x.$$

- Probability:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}.$$

- Algorithm design analysis:
  - Cost of Ax is $O(mn)$, dynamic programming.
- Statistics or machine learning:
  - Maximum likelihood, linear regression.

CPS 340 and auditting 540

- There is also an undergrad ML course, CPSC 340:
  - 340: Lower workload, less math, final exam instead of project.
  - 540: objective is for you to design your own ML methods (when necessary).
  - 340 taught by Raymond Ng, who has more teaching experience.

CPS 340 and auditting 540

- There is also an undergrad ML course, CPSC 340:
  - 340: Lower workload, less math, final exam instead of project.
  - 540: objective is for you to design your own ML methods (when necessary).
  - 340 taught by Raymond Ng, who has more teaching experience.
- Auditting, an excellent option:
  - Pass/fail on transcript rather than grade.
  - Attend lectures and do the coding project.
  - Do the assignments when/if you want to (self-marked).
  - Please do this officially:
    `http://students.ubc.ca/enrolment/coursesreg/academic-planning-resources/auditing-courses`

Textbook

- We will use *Machine Learning: A Probabilistic Approach*:
  - Available for purchase on Amazon.
  - On reserve in reading room (ICCS 262).
  - Available online through the library (see webpage).
  - Many typos but covers most of ML.
  - 1% towards assignment mark for typos (in current edition).
- Other relevant texts include:
  - The Elements of Statistical Learning (Hastie et al.).
  - Pattern Recognition and Machine Learning (Bishop).
  - All of Statistics (Wasserman).

Course Content

- A rough overview of topics and timeline:
  - regression, classification, model selection, regularization, kernels and Gaussian processes, convex and stochastic optimization, bootstrapping/boosting and random forests, mixture and latent variable models, missing data, Bayesian inference, graphical models, and deep learning.

Course Content

- A rough overview of topics and timeline:
  - regression, classification, model selection, regularization, kernels and Gaussian processes, convex and stochastic optimization, bootstrapping/boosting and random forests, mixture and latent variable models, missing data, Bayesian inference, graphical models, and deep learning.
- We will not cover:
  - learning theory (see Nick Harvey's course) or topics involving actions (causality, active learning, reinforcement learning).

Grading

- Homeworks: 30%
- Midterm: 30%
- Coding Project: 10%.
- Final Project: 30%

We will also have a quarter-term teaching evaluation.

Homeworks

- There will be 8 homeworks (only top 6 count).
- Written and Matlab programming.
- Due at the start of class.
- The first one is due Wednesday.

Homeworks

- There will be 8 homeworks (only top 6 count).
- Written and Matlab programming.
- Due at the start of class.
- The first one is due Wednesday.
- Peer marking of written part:
  - End of class on due date: pick up someone else's.
  - Hand in graded homework with your next assignment.
  - Receive graded homework the next class.
  - Thursday tutorial: see the TA about marking errors.
- Late assignments marked by the TA with 25% off.

Getting Help

- You should have Matlab through your department.
  - If not, ask for a CS guest account or purchase through the bookstore.
- Tutorials are 3-4 on Thursdays before assignments due.
  - Optional, main purpose is help on assignments.
  - Mohamed may briefly go over relevant background.
- Use Piazza for assignment/course questions.

Getting Help

- You should have Matlab through your department.
  - If not, ask for a CS guest account or purchase through the bookstore.
- Tutorials are 3-4 on Thursdays before assignments due.
  - Optional, main purpose is help on assignments.
  - Mohamed may briefly go over relevant background.
- Use Piazza for assignment/course questions.
- You can work in groups and use any source, but hand in your own homework and acknowledge sources:
  - 'I worked with Jenny on this problem (she did the proof)'.
  - 'I found this inequality on the Wikipedia entry for norms'.
  - 'I found this exercise online and copied the answer'.

Midterm

- The midterm verifies you can do the assignments:
  - In class November 10.
  - Closed book, two-page double-sided 'cheat seet'.

Midterm

- The midterm verifies you can do the assignments:
  - In class November 10.
  - Closed book, two-page double-sided 'cheat seet'.
- There will be no 'tricks' or 'surprises':
  - I'll give a list of things you need to know how to do.
  - Mostly minor variants on assignment questions.
- You must come see me if you miss the exam with a doctor's note or other relevant documentation.

Coding Project

- We will jointly write a new ML package: *matLearn*.
- The (individual) coding project consists of:
  - Add a new ML method to matLearn (I'll provide a list).
  - There will be a standard coding/documentation style.
  - Make a simple demo of its usage (I'll give examples).
- Due November 26.
- Auditors do the coding project, too.

Final Project

- Projects can be done in groups of 1-3.
- Project proposal due October 29 (maximum 3 pages).
- Possible project ideas:
  - Apply ML to a new domain (from your research?).
  - Compare a variety of ML methods across different tasks.
  - Find a way to scale-up an existing method.
  - Participate in a Kaggle competition.
  - Extend or combine ideas we explored in class.
  - Prove a theoretical result.
  - Add a new task and several models to *matLearn*.
- Final report due December 17
  (maximum 6 pages in Latex using NIPS stylefile, additional appendices may include code or proofs, for coding use Matlab or Python).

Lecture Style and Instructor Evaluation

- I feel that I learn/teach better when using the whiteboard.
  - Slows down the lecture.
  - Makes the lecture adaptive.
- About recording:
  - Please do not record without permission.
  - We'll have someone take a picture of the board.
- Topics/Readings will be posted before each class.
  - If you haven't seen the topic before, please do the reading before class.

Lecture Style and Instructor Evaluation

- I feel that I learn/teach better when using the whiteboard.
  - Slows down the lecture.
  - Makes the lecture adaptive.
- About recording:
  - Please do not record without permission.
  - We'll have someone take a picture of the board.
- Topics/Readings will be posted before each class.
  - If you haven't seen the topic before, please do the reading before class.
- September 29, we'll do an unnofficial instructor evaluation.
  - Will let me adapt the lecture/assignment style.