CPSC 340

Machine Learning and Data Mining

- HOME
- LECTURES
- ASSIGNMENTS
- PROJECT
- PYTHON

About the course

The analysis of data (Biological signals, music, images, video, customer reviews, webpages, medical records, software, game logs, social networks, environmental signals, astro-data, neuron spikes, etc.) is an important research and development frontier. This frontier is expanding vastly thanks to new developments in mathematical modelling, algorithms, data management and computing infrastucture. It is having a profound impact not only in science and medicine, but also in e-commerce, marketing, business and society at large. Inference and learning with massive datasets is also the key ingredient of the intelligent machines of the future.

This course will provide an introduction to this exciting growing field. It will teach the basic principles and skills required for analysing data in a principled way: finding statistical patterns, dimensionality reduction, clustering, classification and prediction. Students will also have the opportunity of learning Python, a widely used programming language.

Logistics

Time: Mon Wed Fri 3:00pm-4:00pm

Location: Hugh Dempster Pavilion 110

Instructor: Nando de Freitas (nando@cs)

TAs: Bobak Shahriari (bshahr@cs), Babak Shakibi (bshakibi@cs) and Pouria TalebiFard (ptfard2@cs)

Tutorial 1 (Bobak): Mon 4:00pm-5:00pm (Forest Sciences Centre 1001)

Tutorial 2 (Babak): Wed 9:00am-10:00am (MacLeod 214)

Office hours (Nando): Wed, Fri 4:00pm-5:00pm (ICICS 123)

Online discussion: cpsc340 google group

Grading

Assignments: 30%
Midterm: 30% (Fri Nov 2nd)
Final: 40% (Dec 17, 8:30 am, PHRM 1101)
The instructor reserves the right to change the marking scheme under
reasonable circumstances.

Assignments

Assignments will involve both written and python programming problems.
All assignments are due on the specified date at 3pm. They are to be handed in
at the classroom where the lecture takes place.
Missing the deadline from Friday to Monday costs 40%, from Monday to
Wednesday 30% and from Wednesday to Friday 30%. For example, if someone
misses a Friday deadline and hands in the homework on the subsequent
Wednesday, the penalty is 70%. Late homeworks will be marked only after the
homeworks that were handed in in a timely fashion have been marked. Hence,
expect a delay in receiving your mark if you handed in late.
Messy homeworks will be penalized - it is your responsibility to ensure that the
material is presented in a clear written form. All pseudocode must be handed in.
Please don't forget to add your name and student number. Please staple your
homework.
Academic honesty is important. If you find the answer to the homework on a
website, book, etc. please acknowledge this in the front page of your homework.
You will not be penalized. We like people who acknowledge the source and we
don't mind if you seek help from friends to solve the homework problems.
However you must ensure that you understand what you are doing, otherwise
you're missing on the learning experience. Not doing the homeworks will also
impact your performance on the exams.

RECOMMENDED READING

- My favourite book for this course is the book of Stuart Russell and Peter
  Norvig titled artificial intelligence. Chapter 14 covers probabilistic
  graphical models. Chapter 15 covers HMMs. Chapter 20 talks about
  maximum likelihood, the EM algorithm, learning the parameters of
  graphical models and naive Bayes. Chapter 18 teaches decision trees,
  linear regression, regularization, neural networks and ensemble learning.
- The machine learning book of Hastie, Tibshirani and Friedman is much
  more advanced, but it is also a great resource and it is free online: The

- elements of statistical learning.
- For graphical models and Beta-Bernoulli models, I recommend A Tutorial on Learning with Bayesian Networks David Heckerman.
- Kevin Murphy has compiled a nice page about Bayesian learning.
- Wikipedia tutorial on the: SVD
- The following handout should help you with linear algebra.

MEDIA

- Machine learning video lectures
- Why stats: NYTimes article
- A video lecture about python's package matplotlib
- The machine learning course of Andrew Ng is available in youtube and iTunes. It is strongly recommended.

This course is offered at UBC computer science by Nando de Freitas.