STA 414/2104, Winter 2018: Statistical Methods for Machine Learning and Data Mining
DRAFT Syllabus (as of Jan 2018)

Instructor:

Prof Mark Ebden, Office: SS6027 CLTA
Email: mark [dot] ebden [at] utoronto [dot] ca
Lectures: Mondays 2-5 pm, SS2135
Office Hours: From 12 to 1 pm on Mondays, by appointment (click here)

Teaching Assistant: Tadeu (Ted)

Course webpage: portal.utoronto.ca

Course materials provided on Portal are for the use of students currently enrolled in this course only. Providing course materials to anyone outside of the course is unauthorized use.

Marking Scheme:

- For undergraduate and graduate students
  - 2-hour midterm: 40%
  - 3-hour final exam: 60%

Graduate students will be evaluated at the graduate level according to the University Assessment and Grading Practices Policy.

For the in-class midterm: You can use a nonprogrammable calculator and an 8x11" Crib Sheet, single-sided.
For the final exam: You can use a nonprogrammable calculator and an 8x11" Crib Sheet, double-sided.

Course Outline:
This course covers some of the theory and methodology of statistical aspects of machine learning. The preliminary set of topics to be covered include:

- Linear methods for regression, Bayesian linear regression
- Linear models for classification
- Probabilistic Generative and Discriminative models
- Regularization methods
- Model Comparison, and BIC if time permits
- Neural Networks
- Optimization
- Radial basis function networks
- Kernel Methods, Gaussian processes
- Mixture models and EM algorithm
- Graphical Models and Bayesian Networks
- Variational inference (if time permits)

Prerequisites:

- CSC108H1/CSC120H1/CSC121H1/CSC148H1
- STA302H1/CSC411H1
- STA303H1 (recommended)

Textbooks: There are no required textbooks for this course. The following are optional references:

- Christopher M. Bishop (2006) Pattern Recognition and Machine Learning, Springer
- Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016), Deep Learning, free at www.deeplearningbook.org
- Machine Learning: A Probabilistic Perspective, by Kevin P. Murphy. (U of T library link here)
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009) The Elements of Statistical Learning (U of T library link here)
- David MacKay (2003) Information Theory, Inference, and Learning Algorithms

Marking concerns

Any requests to have marked work re-evaluated must be made in writing within *one week* of the date the work was returned. The request must contain a justification for consideration.

Missed Tests

- If a test is missed for a valid reason, you must submit documentation to the course instructor.
- If a test is missed for a valid medical reason, you must submit an original copy of the University of Toronto Verification of Student Illness or Injury form to your instructor within two weeks of the test.
- The form will only be accepted as valid if the form is filled out according to the instructions on the form.
- Important: The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing the term test. If the form indicates that the degree of incapacitation on academic functioning is negligible or mild then this will *not* be considered a valid medical reason.
- If a test is missed for a valid reason then the final exam will be worth 60% of your final grade.
- Other reasons for missing a test will require prior approval by your instructor. If prior approval is not received for non-medical reasons then you will receive a term test grade of zero.

Computing

In the (not for credit) assignments you will need to write your own programs, debug them, and use them to conduct various experiments, plot curves, etc; you may use any programming language, but Python, R, and Matlab might be preferable. More information on these languages can be found online. See also datacamp.com.

For Python (recommended):

- Anaconda provides an installer for Python and Numpy for Windows, Linux, and Mac
- Numpy tutorial
- Learn X in Y minutes can get you up to speed in Python if you already know other languages

If you wish to use R, it is freely available for download at http://cran.r-project.org for Windows, Mac, and Linux operating systems. *R Studio* is a good integrated development environment to R. It is freely available at www.rstudio.com/products/rstudio/ To use R at UofT then you will need to sign up for a CQUEST account. To get an account and find out more information about using CQUEST go to www.cquest.utoronto.ca

On the test and exam, you will not be asked to understand a particular language's syntax and will not need to provide code.

Calculators

You will need a calculator. Any calculator that has logarithmic functions will be sufficient. Calculators on phones or other devices equipped to communicate with the outside world (for example, through the internet or cellular or satellite phone networks) will not be permitted during the term test and the final exam.

Online Discussion Board

This term you will have the option to use Piazza for class discussion. If you decide not to use Piazza it will not disadvantage you in any way, and will not affect official University outcomes (e.g., grades and learning opportunities). If you choose not to opt-into Piazza then you can ask questions or discuss course material with the instructor or TAs during office hours.

Be sure to read Piazza's Privacy Policy and Terms of Use carefully. They provide for substantial sharing and disclosure of your personal information. If you decide to participate in Piazza, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use.

The Piazza system is highly catered to getting you help quickly and efficiently from classmates, the TA, and the lecturers. Rather than emailing questions to the teaching staff, we encourage you to post your questions on Piazza. To sign up for the discussion forum, click on the link: piazza.com/utoronto.ca/winter2018/sta414lec0101

Additional help

Need extra help with the coursework? Here are some options:

- For continued class discussion and questions outside of class, try posting on the discussion forums. The instructor and TAs will be monitoring them
- You can visit the instructor or teaching assistants during their office hours
- You may choose to join (or create) a Recognized Study Group: www.studygroups.artsci.utoronto.ca
- E-mail should only be used for emergencies or personal matters

How to communicate with your instructor

Questions about course material such as:

- How do I do question 3.7 in the textbook?
- What is standard deviation?
- When is the midterm?

can be posted on the discussion forums. Questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors), if desired.

For private communication, such as "I missed the test because I was ill," e-mail your instructor. Include your full name and student number.

You may post entirely anonymous feedback here. Nobody will receive this except me, and nobody will know who you are.

Academic integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at www.governingcouncil.utoronto.ca/policies/behaveac.htm. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructor.

Accessibility needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible: accessibility.services@utoronto.ca or http://accessibility.utoronto.ca.

Your responsibilities

The classroom sessions for this class are designed to actively engage you in the course material. We hope you'll find them interesting, challenging, fun, and an excellent opportunity to truly learn the material.