STA 414/2104, Statistical Methods for Machine Learning II
Winter 2019: Syllabus

Instructors:

1. David Duvenaud, Office: PT384
   Email: duvenaud@cs.toronto.edu
   Office: Pratt Building, Room 384

2. Murat A. Erdogdu, Office: SS5016b
   Email: erdogdu@cs.toronto.edu
   Office: Sidney Smith Building, Room 5016b

Lectures: This course has two identical sections each week:

1. Monday 2pm-5pm at Wilson Hall -New College Room 1017,

2. Tuesday 7pm-10pm at Sidney Smith Room 2118.

The two instructors won't stick strictly to lecturing in their own sections.

Teaching Assistants: Hannes Bretschneider, Cedric Beaulac

Course webpages: duvenaud.github.io/sta414 and q.utoronto.ca

Marking Scheme:

- For undergraduate and graduate students
  - 3 assignments: 40%
  - 2-hour midterm 20%
  - 3-hour final exam 40%

Graduate students will be evaluated at the graduate level according to the University Assessment and Grading Practices Policy. Graduate students in STA2104 may have some additional questions to do on assignments (these questions will not be required for undergraduates).

Midterm is (tentatively) in the week of Feb 18th. Details will be announced in class and on the class webpage.
You can bring an optional A4 cheat sheet - double-sided.
Final exam is in April TBD. You can bring an optional A4 cheat 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 may include:

- Overview of machine learning
  - Supervised vs unsupervised learning
  - Least squares
  - Overfitting and generalization
  - Effect of regularization
  - Cross validation
- Probabilistic Models
  - Maximum likelihood estimation
  - Some useful distributions
  - Exponential families

- Regression and classification
- Basis function models
- Classification
  - Generative and discriminative models
  - Logistic regression
  - Discriminant analysis
  - Perceptron
  - Multinomial regression
- Optimization and Decision Theory
  - Bias-variance tradeoff
  - Generalization
  - Statistical decision theory
  - Gradient descent
  - Stochastic gradient descent
  - Newton's method
- Unsupervised learning I:
  - Clustering
  - Mixture models
  - EM algorithm
  - Principle component analysis
- Latent variables
  - Graphical Model notation
  - Markov models
  - Hidden Markov models
  - Exact inference
- Fitting large models
  - Automatic differentiation
  - Vectorization
  - Neural Networks
- Approximate inference
  - MCMC
  - Variational Inference
  - Bayesian neural networks
- Reinforcement learning
  - discrete random variables
- Variational autoencoders
  - Nonlinear dimensionality reduction
  - Recognition networks
- Generative Models
  - Generative adversarial networks
  - Normalizing flows

Prerequisite: Either STA302H or CSC411H

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
- 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
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2017) Introduction to Statistical Learning
- David MacKay (2003) Information Theory, Inference, and Learning Algorithms

Assignments

Late policy: If you are traveling, you may email your solution to one of the course staff in advance of the deadline. Ten percent of the homework value will be deducted for each day a homework is late. No credit will be given for homework submitted after solutions have been posted. Exceptions will be made for documented emergencies.

Collaboration policy: After attempting the problems on an individual basis, you may discuss and work together on the homework assignments with up to two classmates. However, you must write your own code and write up your own solutions individually and explicitly name any collaborators at the top of the homework.

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 the University of Toronto Verification of Student Illness or Injury form to your instructor within one week 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 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, and R might be preferable. More information on these languages can be found online. On the test and exam, you will not be asked to understand a particular language's syntax and will not need to provide code in any particular language. You may be asked to provide pseudo-code.

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 affect official University outcomes (e.g., grades). 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.

The Piazza system is catered to getting you help fast 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, go to the link: piazza.com/utoronto.ca/winter2019/sta414

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.
- 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. Use your U of T e-mail account, put "STA414" in the subject, and include your full name and student number.

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.