

**STA 414/2104, Spring 2017: Statistical Methods for Machine Learning and Data Mining**  
**DRAFT Syllabus (as of 9 Jan 2017)**

**Instructors:**

1. David Duvenaud, Office: PT384  
Email: duvenaud [at] cs [dot] toronto [dot] edu  
Lectures: Mondays 2-5pm, EM 119  
Office hours: Mondays 11 am - 12 pm in Pratt Building, Room 384
2. Mark Ebden, Office: SS6026C and PT371  
Email: mark [dot] ebden [at] utoronto [dot] ca  
Lectures: Tuesdays 7-10 pm, SS1071  
Office Hours: Thursdays 3-4 pm in SS6026C, and after each lecture outside the classroom

**Teaching Assistants:** Amanjit Kainth, Chris Cremer, Luhui Gan, and Yang Guan Jian Guo (Tommy)

**Course webpages:** [duvenaud.github.io/sta414](http://duvenaud.github.io/sta414) and [portal.utoronto.ca](http://portal.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).

**In-class midterm is tentatively set for Monday 13 February.** You can use a nonprogrammable calculator and an 8 by 11 inch Crib Sheet - Single-sided. Handwriting only – no typing.

**Final is in April.** You can use a nonprogrammable calculator and an 8 by 11 inch Crib Sheet - Double-sided. Handwriting only – no typing.

**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
- Neural Networks
- Variational Inference
- Optimization
- Radial basis function networks
- Kernel Methods, Gaussian processes, Support Vector Machines
- Mixture models and EM algorithm
- Graphical Models and Bayesian Networks

**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, free at [www.deeplearningbook.org](http://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](#)

## Assignments

The due dates are Friday 3 February, Friday 10 March, and Friday 31 March. In case you make a mistake that you spot later: you may upload up to three (3) submissions per assignment, and only the most recent one is evaluated.

Assignments are due by 22:00 on the due dates and must be submitted electronically on Portal. Please ensure that you not only upload a *draft* copy of your solutions, but click to finalize/submit the work as well. It is strongly recommended that you do not try to upload or submit your assignment at 21:59 or 21:58 on the due date. Based on past experience there is a good chance that your assignment will be marked late. This will not be considered a valid reason for submitting your assignment late. So, if your assignment is submitted at 22:01 then you will automatically lose 10%.

**Late assignments** will be accepted within 48 hours after the due date.

- Late penalty for assignments: 10% will be deducted from assignments submitted within 24 hours after the due date, and 20% will be deducted from assignments submitted between 24 and 48 hours after the due date
- Assignments submitted more than 48 hours after the due date will receive a grade of zero except for cases with documented reasons beyond the student's control. In the case of a medical reason this must be documented using the UofT [Student Illness or injury form](#). If an assignment is missed for a documented reason then the weight for the assignment will be transferred to the weight for the final exam.
- Email submission of assignments will not be accepted. If an assignment is submitted by email then it will NOT be marked.

## 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 (55% for graduate students).
- 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, R, and Matlab might be preferable. More information on these languages can be found online. For example, 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/](http://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](http://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. Take time to understand and be comfortable with what they say. They provide for substantial sharing and disclosure of your personal information held by Piazza, which affects your privacy. 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 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, click on the link: [piazza.com/utoronto.ca/winter2017/sta414](http://piazza.com/utoronto.ca/winter2017/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 utoronto.ca e-mail account to ensure that your message doesn't automatically go to a Junk folder 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](http://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](mailto: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.