# STA 414/2104, Winter 2018: Statistical Methods for Machine Learning and Data Mining

## Section L5101

**Instructor**:

Alex Stringer, Office: **TBA**
Email: alex [dot] stringer [at] mail [dot] utoronto [dot] ca
Lectures: Tuesdays 7-10 pm, BA1190
Office Hours: Mondays 11-1 pm, **TBA**

**Teaching Assistant:** Tadeu Ferreira

**Course webpage:** portal.utoronto.ca

Materials posted on the course website are strictly confidential, and for the use of students enrolled in the course only. Providing these materials to anyone not currently enrolled in the course is a serious academic office. This will be strictly enforced this semester.

**Marking Scheme:**

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

Graduate students will have the same midterm and final as undergraduates, with possible additional and/or alternative questions.

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

**Midterm**

The midterm will be held during lecture time as follows:

- Date: Tuesday, February 13th, 2018, 7:00 - 9:00 pm
- Location: **TBA**

**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
Neural Networks
Optimization
Kernel Methods, Gaussian processes
Mixture models and EM algorithm
Variational inference

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

Prerequisites will be *strictly* enforced for undergraduate students. Do not approach your instructor to ask to sign a form waiving prerequisites. Undergraduate students without the appropriate prerequisites will be removed from the course. Undergraduate students who do not have exactly the prerequisites listed above, but who have reviewed in detail the contents of the above courses and feel that they have covered this material at a similar level in other courses, may submit a brief description of why they think they should be admitted to the course to ugchair.stats@utstat.utoronto.ca. Include in the subject line "STA414: Request for prerequisite exemption", and briefly explain your situation. Only do this if you have actually reviewed the prerequisites in detail, and are prepared to say why you should be allowed to take the course.

Students who have taken CSC411 may find the material in this course to be very similar to what they have learned already, and may wish to consider taking CSC412, which covers more advanced material.

**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. If you think that a mark has been assigned *in error*, please contact your TA directly via email. Include in the subject line "STA414: Request for Marking Correction", and describe in detail the potential error. The TA will review only the potential error, and adjust your mark upward if warranted. *You must submit this request within one week of the assignment being returned, or it will be ignored.* Requests that do not include the appropriate subject line in the email will be ignored.

If you think your work was marked *unfairly*, please email your instructor with the subject line "STA414: Request for Marking Reconsideration". Include a detailed description of why you think you deserve a higher mark, including explicit reference to any solutions provided. I will decide whether your request is reasonable and if it is, I will remark your paper, and adjust your mark up, down, or not at all. You must submit such requests within 1 week of having the work returned for the final time, i.e. if you are submitting to the TA for a correction and then to me for a re-mark, you must submit to me within one week of the TA returning the assignment to you the second time. The mark I assign after reviewing is final, and non-negotiable. By submitting this request to me, you acknowledge this policy, you accept the mark I give you as your final mark for the item in question, and you waive your right to request any further consideration of your mark for the item in question. Further, by waiting for longer than one week without submitting such a request, you accept the mark assigned as your final grade for the item in question, and waive your right to request further consideration of your mark for the item in question. Requests that do not include the appropriate subject line in the email will be ignored.

## 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: https://piazza.com/utoronto.ca/winter2018/sta414lec5101

## 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.

Questions such as these may also be discussed in the TA's and/or instructor's office hours.

For private communication, such as "I missed the test because I was ill," e-mail your instructor. Include your full name and student number. The subject line of the email should read "STA414: Private Communication", and the body of the email should include your full name and student number as the first two lines.

## 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.

**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 weight of the test will be added to the final exam.

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. 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/ 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.