

# STA314: STATISTICAL METHODS FOR MACHINE LEARNING 1 (FALL 2020/21)

Instructor: Daniel Simpson

**Office Hours:** Monday 4-5pm  
Tuesday 4-5pm  
Other times by appointment only  
TA Office Hours from Week 3 TBA

**Email:** [simpson@utstat.toronto.edu](mailto:simpson@utstat.toronto.edu)

## **Communication:**

In general, I am not able to answer questions about the course material by e-mail. Before you send an e-mail, make sure that you are not asking for information that is already on the course web site, or questions about the course material or assignments that are more appropriately discussed in tutorial or during office hours. If you do not get a response, this may be why. Questions about the course material can be posted on the class Piazza board. This will be monitored by myself and the TAs.

E-mail is appropriate for private communication. Use your utoronto.ca account to ensure that your message doesn't automatically go to my Junk folder. I will generally answer e-mail within one business day.

Announcements will be posted on Blackboard. Please check there regularly. If an urgent matter arises, I may contact the entire class by e-mail. In order to receive these messages, please make sure that your ROSI account has your utoronto.ca e-mail.

## **Lectures:**

The course videos and associated material will appear on Thursdays.

**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: [disability.services@utoronto.ca](mailto:disability.services@utoronto.ca) or <http://studentlife.utoronto.ca/accessibility>.

## Assessment:

	WEIGHT	DATE	TIME (TORONTO)
<b>HOMEWORK 1</b>	5%	Friday 25 September	Due 12:00pm (Midday)
<b>HOMEWORK 2</b>	5%	Friday 2 October	Due 12:00pm (Midday)
<b>ASSIGNMENT 1</b>	20%	Friday 16 October	Due 12:00pm (Midday)
<b>QUIZ 1</b>	20%	Friday 23 October	Open 10:00am–Midnight
<b>HOMEWORK 3</b>	5%	Friday 30 October	Due 12:00pm (Midday)
<b>HOMEWORK 4</b>	5%	Friday 6 November	Due 12:00pm (Midday)
<b>ASSIGNMENT 2</b>	20%	Friday 27 November	Due 12:00pm (Midday)
<b>QUIZ 2</b>	20%	Friday 4 December	Open 10am–Midnight

The four Homeworks will be single question assignments that will be given out a week before it is due. They should be submitted via Quercus.

The two larger Assignments will be submitted through Github (this will be covered in class). These will involve a combination of R programming, data analysis, and writing.

The two Quizzes will be run through Quercus and will be open for a 14 hour period. Students must complete the quiz *in one sitting* and will not be able to repeat questions or restart the quiz. The questions will be randomized from a bank of similar questions.

If a Quiz is missed for a valid reason, you must provide appropriate documentation, such as the University of Toronto Medical Certificate, University of Toronto Health Services Form, or College Registrar's Letter. You must submit this documentation within one week of the test. If documentation is not received in time, your test mark will be zero. If a test is missed for a valid reason, its weight will be shifted to the other assessment.

### Extension policy:

2020 is a challenging year for all of us and, in light of that, this course will have a generous (but not endless) extension policy. It will typically not be difficult to get a one week extension as long as it is requested **at least one day before** the due date. I will only grant longer extensions in extreme and exceptional circumstances, but I am happy to talk to any student about their assessment schedule and I will be as flexible as I can be within the constraints at hand.

Extensions are not cumulative and this unit is frequently assessed so this may lead to you having multiple assessments due on the same day.

*If you have accessibility requirements that involve more time to do assessments please let me know via email early in the semester and we can come up with a workable assessment plan.*

**Re-grading policy:**

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

Regrading requests should only be made for genuine grading errors, and should be initiated by writing or typing a complete explanation of your concern (together with your full name, student number, and e-mail address) on a separate piece of paper, and giving this together with your original unaltered homework/test paper to the instructor within two weeks of when the graded item was first available. Warning: your mark may end up going down rather than up.

**Textbook and slides:**

- This course does not have a text book, but references to appropriate material will be given on the Quercus site.
- Further information will be contained in slides, handouts, and specific references that will be available on Quercus before classes.

**Computing:**

- The course will be run using the R computing environment and will use RStudio and RMarkdown extensively.
- The course will use **github** (<https://github.com>) and all students **must** have a Github account.
- You are strongly encouraged to use RStudio (<https://www.rstudio.com>), which is a free IDE for R.
- All instructions in the course will assume that you have the **latest version** of both RStudio and R installed. We will not answer any R related questions unless both of these things are true.
- The course will make extensive use of the packages associated with the tidyverse (<https://www.tidyverse.org>) and tidymodels (<https://www.tidymodels.org>) projects.

**Course outline:**

This course will cover practical and theoretical aspects relevant to Machine Learning. The course will cover some of the following topics:

- Clustering
- Principal Component Analysis
- Text-as-data
- Linear regression, penalized regression
- Variable selection and the LASSO
- Classification
- Logistic regression
- Decision trees