



DATA*6100 Introduction to Data Science

Fall 2021

0 COURSE PREAMBLE

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings and academic schedules. Any such changes will be announced via Courselink and/or class email. This includes on-campus scheduling during the semester, midterms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website and circulated by email.

Mental Health. One out of every five students in Canada experiences some sort of mental health issue at some point in their academic career. If you find yourself facing a mental health crisis, or just need to talk to someone, please consider taking advantage of one of the following resources available to University of Guelph students:

Counseling Services: (x53244) is located at Health Services (J.T. Powell Building) and offers individual and group counselling sessions by appointment or walk-in.

Student Support Network is located in Raithby House (across from the cannon) and offers confidential, peer-based, drop-in support.

Good2Talk: (1-866-925-5454) is a free, 24/7 student hotline that provides professional counselling and referrals for mental health, addictions and well-being.

Here 24/7: (1-844-437-3247) specializes in assessment, referral and appointment booking and is available 24/7 for crisis support.

You are not alone and you will not be judged for asking for help.

For information on current safety protocols, follow these links:

<https://news.uoguelph.ca/return-to-campus/how-u-of-g-is-preparing-for-your-safe-return/>
<https://news.uoguelph.ca/return-to-campus/spaces/#ClassroomSpaces>

Please note, these guidelines may be updated as required in response to evolving University, Public Health or government directives. Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g. final exam or major assignment).

1 INSTRUCTOR

Instructor: Ayesha Ali

Office: MACN 509 but online for purposes of this course

Phone extension: 53896

E-mail: aali@uoguelph.ca (best way to reach me)

2 AIMS & OBJECTIVES

2.1 Calendar Description

The course includes an introduction to the methods of modern statistics such as splines, general additive models, principal components analysis, classifiers, and Markov random fields. Students learn resampling methods such as bootstrap, cross-validation, boosting, and bagging. Methods of model selection include search-and-score and regularization. Emphasis is on communicating technical ideas to a non-technical audience, including via data visualization.

2.2 Course Description

This course begins with an overview of the data science process from data collection to knowledge transfer and a review of multiple linear regression and link functions. Core data science concepts such as model building, training, model validation, evaluation and statistical testing are discussed with an introduction to algorithmic thinking. Predictive regression models, such as logistic regression, regression splines and generalized additive models are introduced as well as resampling techniques for parameter estimation and cross-validation. State-of-the-art methods for assessing model overfit and performing data reduction through principal components and regularization will be explored. The concepts of Markov random fields and restricted Boltzmann machines are then introduced while the course ends with a look at non-parametric statistical tests.

Course topics include:

- Overview of data analysis and data ethics; review of multiple linear regression
- Fundamentals of model building: data splitting for training, validating and testing
- Nonlinear regression models: logistic regression, regression splines, generalized additive models; estimation versus prediction
- Resampling methods: bootstrapping, permutations, boosting, bagging, parallel computing
- Data reduction through principal components, model scoring, variable selection and regularization
- Markov random fields, Ising models, restricted Boltzmann machines
- Non-parametric tests based on resampling methods

Prerequisite(s): Enrollment in the data science program or permission of instructor.

Credit Weight: 0.5

Academic Department: Mathematics & Statistics

2.3 Learning Outcomes

Upon successful completion of this course, students will have demonstrated the ability to:

1. Understand full data analysis process: data collection, cleaning, description, exploratory data analysis, model building, model checking, model selection, knowledge transfer.
2. Perform data splitting for model training, model validation and model testing; differentiate estimation from prediction task.
3. Demonstrate understanding of regression model framework; differences between linear, generalized linear and generalized additive models.
4. Build nonlinear regression model based on logistic regression, regression splines and generalized additive models.
5. Perform resampling techniques, such as bootstrapping, cross-validation, boosting, and bagging.
6. Perform data reduction through principal components, model scoring and variable selection.
7. Demonstrate basic understanding of Markov random fields and Ising models.
8. Design non-parametric test based on permutations or bootstrapping.
9. Understand basics of data ethics: fairness, bias, privacy, transparency, security, impact and consequences.

3 TEACHING AND LEARNING ACTIVITIES

3.1 Timetable

Lectures: Synchronous lectures via Courouselink MW 4:00-5:20pm EST

Office Hours : TBD and provided on Courouselink once set. Office hours may change.

Due to the Fall Study Break, lecture for Monday, October 11 is rescheduled for Friday December 3

3.2 Course Topics and Schedule

Week	Topic
Week 1	Overview of data analysis and data ethics: fairness/bias, privacy/transparency, security, impact/consequences
Week 2-3	Review of multiple regression, Nonlinear regression
Weeks 3-4	Data splitting, model building, estimation vs prediction
Weeks 5-7	Resampling techniques
Weeks 8-9	Data reduction
Week 10	Markov random fields, Ising models, restricted Boltzmann machines
Weeks 11-12	Non-parametric tests

4 LEARNING RESOURCES

4.1 Course Website

Course material, news, announcements, and grades will be regularly posted to the DATA*6100 Courouselink site. You are responsible for checking the site regularly.

4.2 Readings and Resources

Required Text: None

Recommended Texts:

1. Gareth James, Daniela Whitten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*, 8th Ed. Springer, 2021.
2. Mariarosaria Taddeo and Luciano Floridi. *The ethical impact of data science*, Thematic Issue of Philosophical Transactions of the Royal Society A, Mathematical, Physical and Engineering Sciences, Vol 374, Issue 2083, December 2016.
3. Other relevant readings may/will be provided throughout the semester.

5 ASSESSMENT

5.1 Dates and Distribution

Assignment	Due Date	Weighting	Learning Outcome(s) Assessed
Assignment 1	Week 3	15%	1,2,8
Assignment 2	Week 6	15%	2,3,4,8
Case Study 1	Weeks 6-8	10%	2,5,8
Case Study 2	Weeks 9-11	10%	2,6,8
Project – Proposal	Week 9	5%	1-7
Project – Overview, preliminaries	Week 12	7%	1-7
Project – Final project submission	Week 14	38%	1-8

5.2 Assessment Descriptions

Assessments will require students to discuss the data acquisition, analysis method, and/or policy impact of the analysis from a data ethics perspective, within the context of the scientific question to be addressed by the analysis. **Assignments** provide hands-on experience with analyzing data using relevant software and critically evaluating model performance while **case studies** provide opportunities for group work, literature review, analysis, and presentation. The **final project** will allow students to apply the techniques learned in the course to real world problems, but with clear stop points for feedback, through the **proposal** and **project overview** assessments. **Final projects** will be in the form of formal statistical reports that include a section on data ethics. **Late assignments** not accompanied by a reasonable medical or personal excuse are docked 5% per day late and are not accepted more than a week late.

Assignments and project proposal and overview/preliminaries will be due on Thursdays at 11:59pm EST in weeks 3, 6, 9 and 12, respectively. Case Study presentations will be scheduled for during class time in weeks 8 and 11, with specific dates and times depending on the number of groups. Final presentation slides must be submitted by 11:59pm EST on Wednesdays in weeks 8 and 11. The final project will be due Monday, December 13, 2021 at 11:59pm.

5.3 Course Grading Policies

Passing grade: In order to pass the course, students must obtain a grade of 65% or higher on the total mark of all assessments.

Course Policy on Group Work: Students are encouraged to work together but must each write up individually the material they turn in. Work on the final project may be done in consultation with other students or with faculty but the final project must be completely written by the student.

Accommodation of Religious Obligations: If you are unable to meet an in-course requirement due to religious obligations, please email the course instructor within two weeks of the start of the semester to make alternate arrangements. See the graduate calendar for information on regulations and procedures for Academic Accommodation of Religious Obligations: [\[https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2228.shtml\]](https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2228.shtml)

6 UNIVERSITY STATEMENTS

6.1 Email Communication

As per university regulations, all students are required to check their e-mail account regularly; e-mail is the official route of communication between the University and its students.

6.2 When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons, please advise the course instructor in writing, with your name, id#, and e-mail contact. See the academic calendars for information on regulations and procedures for Academic Consideration:

Undergraduate Calendar - Academic Consideration and Appeals

[\[https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml\]](https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml)

Graduate Calendar - Grounds for Academic Consideration

[\[https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2182.shtml\]](https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2182.shtml)

6.3 Drop Date

Courses that are one semester long must be dropped by the end of the last class day; two-semester courses must be dropped by the last day of class in the second semester. The regulations and procedures for changing graduate course registration are available in the Undergraduate and Graduate Calendars.

Undergraduate Calendar – Dropping Courses

[\[https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml\]](https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml)

Graduate Calendar – Registration Changes

[\[https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml\]](https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml)

6.4 Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments; you may be asked to resubmit work at any time.

6.5 Accessibility

The University of Guelph is committed to creating a barrier-free environment. Providing services for students is a shared responsibility among students, faculty and administrators. This

relationship is based on respect of individual rights, the dignity of the individual and the University community's shared commitment to an open and supportive learning environment. Students requiring service or accommodation, whether due to an identified, ongoing disability or a short-term disability should contact the Student Accessibility Services (SAS) as soon as possible. Documentation to substantiate the existence of a disability is required; however, interim accommodations may be possible while that process is underway. It should be noted that common illnesses such as a cold or the flu do not constitute a disability. For more information, see SAS the website: [\[http://www.uoguelph.ca/sas\]](http://www.uoguelph.ca/sas)

6.6 Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community—faculty, staff, and students—to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar – Academic Misconduct

[\[https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml\]](https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml)

Graduate Calendar – Academic Misconduct

[\[https://www.uoguelph.ca/registrar/calendars/graduate/2018-2019/genreg/sec_d0e2632.shtml\]](https://www.uoguelph.ca/registrar/calendars/graduate/2018-2019/genreg/sec_d0e2632.shtml)

Inappropriate online behaviour will not be tolerated. Examples of inappropriate online behaviour include:

- Posting inflammatory messages about your instructor or fellow students
- Using obscene or offensive language online
- Copying or presenting someone else's work as your own
- Adapting information from the Internet without using proper citations or references
- Buying or selling term papers or assignments
- Posting or selling course materials to course notes websites
- Having someone else complete your quiz or completing a quiz for/with another student
- Making false claims about lost quiz answers or other assignment submissions
- Threatening or harassing a student or instructor online
- Discriminating against fellow students, instructors or TAs
- Using the course website to promote profit-driven products or services
- Attempting to compromise the security or functionality of the learning management system

- Sharing your username and password
- Recording lectures without the permission of the instructor

6.7 Recording of Materials

Presentations that are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

6.8 Other University Resources

The Academic Calendars [<https://www.uoguelph.ca/academics/calendars>] are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.