

SS 4850G/SS 9850 (Advanced Data Analysis) Winter 2021 Course Syllabus

1. Course Information

Instructor	Day/Time	Delivery mode and location*	Contact	Office hours
Dr. Camila de Souza	Mondays, Wednesdays, and Fridays 12:30-1:30 pm (EST)	<ul style="list-style-type: none"> • Synchronously online via zoom until January 31st • In person at PAB 117 after January 31st 	use OWL messages (contact “Instructor Role”) – not UWO email	Wednesdays at 2 pm online via zoom until Jan 31st and at WSC 225 after Jan 31st

*Although the intent is for this course to be delivered in-person to the extent possible, the changing COVID-19 landscape may necessitate some or all of the course to be delivered online, either synchronously (i.e., at the times indicated in the timetable) or asynchronously (e.g., posted on OWL for students to view at their convenience), as deemed most appropriate by the instructor. The grading scheme will not change. Any assessments affected will be conducted online as determined by the course instructor.

Description:

Modern methods of data analysis including ridge regression, lasso, linear discriminant analysis, nonparametric regression, bootstrap, EM algorithm, Bayesian inference, classification trees, neural networks, dimensionality reduction techniques, and clustering.

→ See page 7 for tentative course schedule.

Prerequisite(s):

- A minimum mark of 60% in both Statistical Sciences 3843A/B and Statistical Sciences 3859A/B.
- Knowledge of R/RStudio including the use of markdown to produce beautiful PDF technical reports is assumed or that you are willing to learn.

Unless you have either the requisites for this course or written special permission from your Dean to enroll in it, you may be removed from this course and it will be deleted from your record. This decision may not be appealed. You will receive no adjustment to your fees in the event that you are dropped from a course for failing to have the necessary prerequisites.

2. Communication

To communicate with the instructor **always use the OWL messages tool (to “Instructor Role”)**. Messages sent to instructor’s UWO email will **NOT** be replied.

You can expect a response to a message to the instructor within **approximately 48 hours** during the work week (during busy times, it may take a little longer). **Note that messages will not be answered within the 24-hour period before exams or project deadlines**, nor can I guarantee responses over weekends/holidays.

Students should check OWL (<http://owl.uwo.ca>) on a regular basis for news and updates. This is the primary method by which information will be disseminated to all students in the class. Students are responsible for checking OWL on a regular basis.

If students need assistance with the course OWL site, they can seek support on the OWL Help page. Alternatively, they can contact the Western Technology Services Helpdesk. They can be contacted by phone at 519-661-3800 or ext. 83800.

3. Course Materials

- A laptop or computer
- A working microphone and webcam
- A reliable internet connection

Chapters and sections from the following books (all available online at UWO library):

- The Elements of Statistical learning – also available at author’s webpage:
<https://web.stanford.edu/~hastie/ElemStatLearn//>
- An Introduction to Statistical Learning with Applications in R – also available at author’s webpage: <https://trevorhastie.github.io/ISLR/>
- Kuhn, Max, and Kjell Johnson. *Applied predictive modeling*. New York: Springer, 2013
- Hoff, Peter D. *A first course in Bayesian statistical methods*. New York: Springer, 2009

R statistical software: This course is heavily based on R and, therefore, all assignments and reports will require **coding in R and Rmarkdown**. Please make sure you have the latest R version installed in our computer (<https://cran.r-project.org/>) as well as R studio (<https://www.rstudio.com/products/rstudio/download/>).

4. Methods of Evaluation

Component	Weight	Deadlines/Due dates
Assignment 1	5%	Friday February 4
Assignment 2	7.5%	Friday March 11
Assignment 3	7.5%	Friday April 8

Midterm: <u>individual</u> report on a research paper involving advanced data analysis	35%	Friday March 4
Final Exam: data analysis project report (<u>individual or in pairs</u>)	45% (10% for first part and 35% for complete final report)	<ul style="list-style-type: none"> First project part (dataset description and methodology plan) is due Friday Mar 25 Complete final project report is due Wednesday April 20

i) Assignments

Assignments will be available on the course OWL site. However, you will not submit your solutions to OWL. Instead, **assignments must be submitted through Gradescope** (<https://www.gradescope.com/>) an on-line collaborative grading system. It is your responsibility to make sure that your assignment is successfully uploaded and legible. Submissions that cannot be read by the grader will receive a grade of zero.

Assignment submissions are due 11:55pm (Eastern Time) on the due date. No credit will be given for submissions beyond this time unless a valid academic accommodation is obtained (see Section 6 for details on accommodation). **Students obtaining such appropriate accommodation will have to submit their assignment within 24 hours of the end of the accommodation period.** Failure to submit a missed assignment within 24 hours of a valid academic accommodation period will result in a grade of 0%.

ii) Midterm

The midterm consists of a written summary report on a research paper that used advanced data analysis methods. Each student will work on a different paper. More details will be given in the first week of classes. The report is **due 11:55pm (Eastern Time) on the due date** and **submissions must be conducted via Gradescope**. No credit will be given for submissions beyond this time unless a valid academic accommodation is obtained (see Section 6 for details on accommodation). Those students obtaining appropriate accommodation must contact the instructor immediately to set a new deadline.

iii) Final exam

The final exam consists of a project report which can be conducted individually or in pairs (two students). The project involves a comprehensive advanced statistical analysis of a dataset. Each student (or pair of students) will work on different dataset. More details will be given in the first week of classes. Submissions of the first part and complete report are **due 11:55pm (Eastern Time) on the due dates and must be conducted via Gradescope**. No credit will be given for submissions beyond due dates unless a valid academic accommodation is obtained (see Section 6 for details on accommodation). Those students obtaining appropriate accommodation must contact the instructor immediately to set a new deadline.

iv) Rounding of marks

Across the Sciences Undergraduate Education programs, we strive to maintain high standards that reflect the effort that both students and faculty put into the teaching and learning experience during this course. All students will be treated equally and evaluated based only on their actual achievement.

Final grades on this course, irrespective of the number of decimal places used in marking individual assignments and tests, will be calculated to one decimal place and rounded to the nearest integer, e.g., 74.4 becomes 74, and 74.5 becomes 75. Marks WILL NOT be bumped to the next grade or GPA, e.g. a 79 will NOT be bumped up to an 80, an 84 WILL NOT be bumped up to an 85, etc. The mark attained is the mark you achieved, and the mark assigned; requests for mark “bumping” will be denied.

5. Student Absences

Academic Consideration for Student Absences

Students who experience an extenuating circumstance (illness, injury or other extenuating circumstance) sufficiently significant to temporarily render them unable to meet academic requirements may submit a request for academic consideration through the following routes:

- (i) Submitting a Self-Reported Absence (SRA) form provided that the conditions for submission are met (applicable only to undergraduate students). To be eligible for a Self-Reported Absence:
 - an absence must be no more than 48 hours
 - the assessments must be worth no more than 30% of the student’s final grade
 - no more than two SRAs may be submitted during the Fall/Winter term
- (ii) For medical absences, submitting a Student Medical Certificate (SMC) signed by a licensed medical or mental health practitioner to the Academic Counselling office of their Faculty of Registration.
- (iii) Submitting appropriate documentation for non-medical absences to the Academic Counselling office in their Faculty of Registration.

Students should also note that individual instructors are not permitted to receive documentation directly from a student, whether in support of an application for consideration on medical grounds, or for other reasons. **All documentation required for absences that are not covered by the Self-Reported Absence Policy must be submitted to the Academic Counselling office of a student's Home Faculty.**

For the policy on Academic Consideration for Student Absences – Undergraduate Students in First Entry Programs, see:

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_illness.pdf

and for the Student Medical Certificate (SMC), see:

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/medicalform.pdf.

Religious Accommodation

When a course requirement conflicts with a religious holiday that requires an absence from the University or prohibits certain activities, students should request accommodation for their absence in writing at least two weeks prior to the holiday to the course instructor and/or the Academic Counselling

office of their Faculty of Registration. Please consult University's list of recognized religious holidays (updated annually) at

<https://multiculturalcalendar.com/ecal/index.php?s=c-univwo>.

6. Accommodation and Accessibility

Accommodation Policies

Students with disabilities work with Accessible Education (formerly SSD), which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The policy on Academic Accommodation for Students with Disabilities can be found at:

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Accommodation_disabilities.pdf,

7. Academic Policies

The website for Registrarial Services is <http://www.registrar.uwo.ca>.

In accordance with policy,

https://www.uwo.ca/univsec/pdf/policies_procedures/section1/mapp113.pdf,

the centrally administered e-mail account provided to students will be considered the individual's official university e-mail address. It is the responsibility of the account holder to ensure that e-mail received from the University at his/her official university address is attended to in a timely manner.

Scholastic offences are taken seriously, and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site:

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf.

All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).

In addition, online services such as Chegg are actively monitored. Any questions that are coming out from assignments and are posted to an online service will be searched. Such an activity will be considered as a scholastic offence and will result in academic penalty.

Participants in this course are **not permitted to record the sessions**, except where recording is an approved accommodation, or the participant has the prior written permission of the instructor.

Please be aware that all course materials created by the instructor(s) are copyrighted and **cannot be sold/shared**. Those include materials used in tests/quizzes, midterms, and finals. Any posting/sharing of

such materials in part or whole without owner's consent is considered as violation of the Copyright Act and will be considered as a scholastic offence.

8. Support Services

Please visit the Science & Basic Medical Sciences Academic Counselling webpage for information on adding/dropping courses, academic considerations for absences, appeals, exam conflicts, and many other academic related matters: <https://www.uwo.ca/sci/counselling/>.

Please contact the course instructor if you require lecture or printed material in an alternate format or if any other arrangements can make this course more accessible to you. You may also wish to contact Accessible Education at (519) 661-2147 if you have any questions regarding accommodations.

Learning-skills counsellors at the Student Development Centre (<http://www.sdc.uwo.ca>) are ready to help you improve your learning skills. They offer presentations on strategies for improving time management, multiple-choice exam preparation/writing, textbook reading, and more. Individual support is offered throughout the Fall/Winter terms in the drop-in Learning Help Centre, and year-round through individual counselling.

Students who are in emotional/mental distress should refer to Mental Health@Western (<http://www.health.uwo.ca/mentalhealth>) for a complete list of options about how to obtain help.

Additional student-run support services are offered by the USC, <http://westernusc.ca/services>.

9. Tentative course schedule

Week	Topics	Reminders
1	Jan 10-14	- Statistical learning - Supervised vs unsupervised learning
2	Jan 17-21	Supervised Learning - Ridge regression and LASSO
3	Jan 24-28	- Classification: Discriminant Analysis, Logistic regression, Naïve Bayes, kNN - ROC
4	Jan 31-Feb 4	- Mitigating class imbalance
5	Feb 7-11	- Nonparametric regression
6	Feb 14-18	- Model assessment
7	Feb 21-25	Reading week
8	Feb 28-Mar 4	- Overview of tree-based methods and neural networks
9	Mar 7-11	Inference - Bootstrap versus maximum likelihood - Expectation-Maximization algorithm

10	Mar 14-18	- Introduction to Bayesian Inference	
11	Mar 21-25	- More on Bayesian Inference	First part of final project is due Friday Mar 25
12	Mar 28-Apr 1	Unsupervised Learning - Clustering: K-means, hierarchical clustering	
13	Apr 4-8	- Dimensionality reduction techniques: PCA NMF, MDS April 8th is the last day of classes	Assignment 3 is due Friday Apr 8