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

Although this academic year might be different, Western University is committed to a **thriving campus**. We encourage you to check out the Digital Student Experience website to manage your academics and well-being. Additionally, the following link provides available resources to support students on and off campus: https://www.uwo.ca/health/.

Technical Requirements and Important Dates:


Stable internet connection


Laptop or computer


Working microphone


Working webcam

Classes Start	Reading Week	Classes End	Study day(s)	Exam Period
January 11	February 15 - 19	April 12	April 13	April 14 - 30

* March 15, 2021: Last day to drop a second-term half course or a second-term full course without penalty

1. Course Information

Instructor	Day/Time	Location	Contact	Office hours
Camila de Souza	Mondays, Wednesdays, and Fridays 12:30-1:30 pm (EST)	Synchronous Online via Zoom	use OWL messages (contact "Instructor Role") – **not** UWO email	Tuesdays 1 to 2 pm (EST) online via Zoom

Description:
Modern methods of data analysis including ridge regression, Lasso, linear discriminant analysis, nonparametric regression, bootstrap, EM algorithm, classification trees, neural networks, dimensionality reduction techniques, and clustering. If time allows additional topics may be taught.

→ See page 7 for the course schedule.

Prerequisite(s):

- A minimum mark of 60% in both Statistical Sciences 3843A/B and Statistical Sciences 3859A/B.

- A knowledge of R/RStudio including the use of markdown to produce beautiful PDF technical reports is assumed or that you are willing to learn.

2. Communication

To communicate with the instructor **always use the OWL messages (to "Instructor Role")** tool. 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.

The Forum tool is enabled on the OWL website. Please use this Forum to post and respond to questions about course content (e.g., lecture, readings, etc). The Forum will be monitored on a regular basis and the instructor or TAs will interject with corrections or responses as necessary. As this is an open Forum, please be respectful of your peers, instructor(s), and TAs. Derogatory, discriminatory, or otherwise inappropriate language or topics will be removed and dealt with at the instructor's discretion.

3. Online Participation and Engagement

☒ Attendance at synchronous sessions is required
☒ Sessions will **not** be recorded but all slides will be available on OWL
☒ Students are expected to participate and engage with content as much as possible
☒ Students can participate during the live class sessions by raising your hand on Zoom or using the chat option to post a question. I plan to stop every once in a while during the live classes to check to see if any student has posted a question or has raised their hand.
☒ Students can also participate by interacting in OWL forums with their peers, instructor and TAs.

4. Course Materials

Required texts:

1. The Elements of Statistical learning (**ESL**) – pdf freely available at UWO library or author's webpage: https://web.stanford.edu/~hastie/ElemStatLearn//

2. An Introduction to Statistical Learning with Applications in R (**ISLR**) - pdf freely available at UWO library or author's webpage: https://trevorhastie.github.io/ISLR/

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/).

5. Methods of Evaluation

Component	Weight	Deadlines/Due dates
Assignment 1	6.25%	Friday Jan 29
Assignment 2	6.25%	Friday Feb 26
Assignment 3	6.25%	Friday March 26
Assignment 4	6.25%	Monday April 12
Midterm: individual report on a research paper involving advanced data analysis	30%	Friday March 5
Final Exam: data analysis project report (in groups)	45%	Friday April 23

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).

Solutions to assignments **will not** be posted; however, TAs will provide comments on wrong answers using Gradescope, which will allow students to find out the correct solutions. In addition, students can ask the instructor and TAs for more details on solutions during office hours.

ii) Midterm

The midterm consists of a written summary report on a research paper that used advanced data analysis methods and it is **due 11:55pm (Eastern Time) on the due date**. **Submissions must be conducted via Gradescope**. Each student will work on a different paper. More details will be presented in the first weeks of classes.

iii) Final exam

The final exam consists of a group project report. Students will be **randomly assigned** to groups via an R script that will be run live during class. The project involves a comprehensive advanced statistical analysis of a dataset. Each group will work on a different dataset. Submission is **due 11:55pm (Eastern Time) on the due date and must be conducted via Gradescope**. More details will be presented in the first weeks of classes.

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 reports, 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.

6. Accommodation and accessibility

- If you are unable to submit an assignment by the due date/time, you must seek approval for the absence as soon as possible and contact your instructor immediately. Approval can be granted either through a self-reporting of absence or via the Dean's Office/Academic Counselling unit of your Home Faculty. Those **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%.

- If you are unable to submit your midterm report by the due date/time, you must seek approval for the absence as soon as possible and contact your instructor immediately. Approval can be granted either through a self-reporting of absence or via the Dean's Office/Academic Counselling unit of your Home Faculty. Those **students obtaining such appropriate accommodation will have to submit their midterm report within 24 hours of the end of the accommodation period** otherwise their grade will be 0%.

- **There is no accommodation for the final project report**, since one of the group members should be able to submit the report by the due time and date.

- Students with disabilities work with Accessible Education (formerly SSD) which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The Academic Accommodation for Students with Disabilities policy can be found at: https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic Accommodation_disabilities.pdf

- Religious Accommodation: students should consult the University's list of recognized religious holidays, and should give reasonable notice in writing, prior to the holiday, to the Instructor and an Academic Counsellor if their course requirements will be affected by a religious observance. Additional information is given in the Western Multicultural Calendar: https://multiculturalcalendar.com/ecal/index.php?s=c-univwo

7. Academic Policies

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

In accordance with policy, http://www.uwo.ca/its/identity/activatenonstudent.html, 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.

If you are unable to meet a course requirement due to illness or other serious circumstances, you must seek approval for the absence as soon as possible. Approval can be granted either through a self-reporting of absence or via the Dean's Office/Academic Counselling unit of your Home Faculty. For further information, please consult the university's policy on academic consideration for student absences: https://tinyurl.com/AcademicConsiderations

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. 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).

Completion of this course will require you to have a reliable internet connection and a device that meets the system and technical requirements for Zoom. Information about the system and technical requirements are available at the following links:
https://support.zoom.us/hc/en-us

* Please note that Zoom servers are located outside Canada. If you would prefer to use only your first name or a nickname to login to Zoom, please provide this information to the instructor in advance of the test or examination.

Professionalism & Privacy

Western students are expected to follow the Student Code of Conduct. Additionally, the following expectations and professional conduct apply to this course:

☒ Students are expected to follow online etiquette expectations provided on OWL

☒ All course materials created by the instructor(s) are copyrighted and cannot be sold/shared

☒ Recordings are not permitted (audio or video) without explicit permission

☒ Permitted recordings are not to be distributed

☒ Students will be expected to take an academic integrity pledge before some assessments

Copyright Statement

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.

In addition, online services such as Chegg are actively monitored. Any questions that are coming out during midterms and finals 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.

8. Support Services

Please visit the Science & Basic Medical Sciences Academic Counselling webpage for information on add/drop 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 Student Accessibility Services (SAS) at 661-2147 if you have any questions regarding accommodations.

Western University is committed to a thriving campus as we deliver our courses in the mixed model of both virtual and face-to-face formats. We encourage you to check out the Digital Student Experience website to manage your academics and well-being: https://www.uwo.ca/se/digital/.

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/mental_health) 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. Course Schedule

Week		Topics	Text Material	Reminders
1	Jan 11 – 15	- Statistical learning - Supervised vs unsupervised learning	ISLR: 2.1 all the way to the end of 2.2.3	
2	Jan 18 – 22	**Supervised Learning** - Brief overview of linear regression - Ridge regression and LASSO	ESL: 3.4.1, 3.4.2 and 3.4.3	
3	Jan 25 – Jan 29	- Classification: LDA, Logistic regression, Naïve Bayes, kNN - ROC	ISLR: 4.3, 4.4, 4.5, 4.6 ESL: 4.3, 4.3.1, 4.3.2	Assignment 1 Jan 29
4	Feb 1 – 5	- Nonparametric regression	ESL: 5.1, 5.2, 5.4, 5.5, 6.1, 6.2, 6.3	
5	Feb 8 – 12	- Model assessment	ESL: 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7	
6	Feb 15 – 20	**Reading week**		
7	Feb 22 – 26	- Bootstrap versus maximum likelihood - EM algorithm	- ESL: 8.2 - ESL: 8.5.1 and 8.5.2	Assignment 2 Feb 26
8	Mar 1 – 5	- Tree-based methods - Bagging	- ESL: 9.2.1, 9.2.2, 9.2.3, 9.2.5, ISLR: 8.1, 8.3 - ESL 8.7, ISLR 8.2.1, 8.3	Midterm March 5
9	Mar 8 – 12	- Boosting trees - Random forests	- ESL 10.9 and 10.10; ISL 8.2.3, 8.3 - ESL Ch. 15, ISLR 8.2.2, 8.3	
10	Mar 15 – 19	- Neural Networks	ESL: 11.3, 11.4, 11.5, 11.6, 11.7	Assignment 3 Mar 26
11	Mar 22 – 26	**Unsupervised Learning** - Clustering: K-means, hierarchical clustering, SOMs	ESL 14.3, 14.4	
	March 29 – April 2	- Dimensionality reduction techniques: PCA (ESL 14.5.1), NMF (ESL 14.6), MDS (ESL 14.8)		
	April 5 -9	- Undirected Graphical Models (if time permitted)	ESL 17.3, 17.4	
	April 12	- Last day of classes		Assignment 4