

1. General Course Information

Course Information

INTRODUCTION TO DATA SCIENCE I
CS4414B/CS9637B/CS9114B
Winter 2021

Lecture: Online Asynchronous – OWL
Office hours: Monday 12:30-1:30 (drop in)
Tutorial/Q&A: Wednesday 2:30pm to 4:30pm on Zoom or Chat
Place: Online

List of Prerequisites

Antirequisite(s): [Computer Science 4437A/B/Y](#) if taken during Fall/Winter 2015 or 2016, SS3650, SE4460A, SS3850B. If you require to take SS3850 and CS4414 for a Major or Honor Specialization in Data Science, please contact me.

Prerequisite(s):

0.5 course from Biology 2244A/B, Statistical Sciences 2035, Statistical Sciences 2141A/B, Statistical Sciences 2143A/B, Statistical Sciences 2244A/B or Statistical Sciences 2858A/B;
1.0 course from Computer Science 1025A/B, Computer Science 1026A/B, Computer Science 1027A/B, Computer Science 1037A/B, Computer Science 2120A/B, Computer Science 2121A/B, Digital Humanities 2220A/B, Digital Humanities 2221A/B, Engineering Science 1036A/B; and
0.5 course from Mathematics 1229A/B, Mathematics 1600A/B, Applied Mathematics 1411A/B; or permission of the Department.

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.

For graduates: There are no set prerequisites for the class. However, a decent working knowledge of probability, linear algebra, and basic programming in Python is required. A more detailed self-assessment on whether students have the required background is available at:

<https://www.csd.uwo.ca/~dlizotte/teaching/IDS/index.html>

2. Instructor Information

Professor: Jörn Diedrichsen
Email: jdiedric@uwo.ca
Office: WIRB 4138 (Mondays 11am-12 – by appointment)

Students must ask question via the OWL forum. In the case of a general question, students should check the OWL forum first for an answer and then post the question to the OWL forum. Email communication will be reserved for strictly personal matters.

3. Course Description/Syllabus

Introduces machine learning and statistical methods for data analysis through applied examples. Particular emphasis is placed on how to rigorously evaluate an analysis of data. Also suitable for non-Computer Science students with appropriate background.

Topics include:

- Supervised Learning and Model Fitting
- Statistics, Prediction, and Maximum Likelihood
- Introduce test set/out-of-sample idea.
- Classification, Evaluation, Logistic regression Regularization, Multi-class problems
- Estimating Performance, Quantifying Uncertainty on parameter estimates and on model predictions
- Test error, Cross-validation, Model Selection, Bias-Variance tradeoff
- Feature Selection and Regularization (L1 and L2)
- Trees, Random Forest
- Neural Networks, Gradients, learning
- Autoencoders, Dimensionality reduction, PCA, NMF, tSNE
- Clustering, K-means, hierarchical clustering
- Model limitations, Causality.

Table of Contents and Schedule

Week	Dates	Lecture	Lab	Assignment
1	Jan 11-15	Supervised Learning and Model Fitting	Pandas and Numpy, Optimization	Linear regression, Loss function Squared error Mean absolute deviation
2	Jan 18 – 22	Probability and Maximum Likelihood	Pandas	Regression
3	Jan 25 – 29	Introduce test set/out-of-sample idea. Classification, Evaluation, Logistic regression Regularization, Multi-class problems	Implementation of topics	Logistic regression, Classification, Evaluation
4	Feb 1 – 5	Estimating Performance, Quantifying Uncertainty	Bootstrap	Bootstrap and confidence intervals
5	Feb 8 – 12	Test error, Cross-validation, Model Selection, Bias-Variance tradeoff	Cross-validation	Cross-validation and model selection
6	Feb 15 – 19	Reading Week	N/A	
7	Feb 22 – 26	Feature Selection and Regularization (L1 and L2)	Regularization	Regularization and nested cross-validation
8	Mar 1 – 5	Midterm		-
9	Mar 8 –12	Trees, Random Forest	Tree Lab	Tree homework
10	Mar 15 – 19	Neural Networks, Gradients, learning	Simple 1-hidden layer network	3-layer network on non-linear problem
11	Mar 22 – 26	Autoencoders, Dimensionality reduction, PCA, NMF, tSNE	Dimensionality reduction	Dimensionality reduction
12	Mar 29 – Apr 2	Clustering, K-means, hierarchical clustering	Clustering	Clustering
13	Apr 5 – 9	Model limitations, Causality.	Repetition of techniques, critical appraisal	Repetition of techniques, critical appraisal

4. Course Materials

Lecture and Tutorial Videos

Lecture and Tutorial videos will be available on OWL.

If students need assistance, 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.

Recommended Texts

HTF: *The Elements of Statistical Learning* by Hastie, Tibshirani and Friedman. [[Free online](#)]

MLPP: *Machine Learning: a Probabilistic Perspective* by Kevin P. Murphy [[Free online](#)]

Students must 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.

5. Methods of Evaluation

The overall course grade will be calculated as listed below:

- Weekly Assignments - 40%
- Midterm - 25%
- Final Exam - 35%

Weekly Assignment:

Assignments will be released each week, with due dates of the following week. No late Assignments will be accepted, but will be graded with 0%. There will be no make-up for missed weekly assignments without any exceptions. If the student submits a self-reported absence before the due date of the assignment, a 48hr extension of the deadline will be granted (see below).

Midterm:

The midterm exam will be held synchronously, most likely Wednesday, March 3rd, 10:30-13:30. The midterm will cover concepts up to and including those covered up to this point.

The midterm will be a practical examination in the form of a timed assignment. Students will be given a data set and a set of practical data analytic problems to solve, similar to the structure of the weekly assignments. The exam is “open book & open web”, meaning that students can access any notes or any documents on the web. Electronic communication with other people inside or outside class is prohibited.

For Undergraduates:

The 3 hr final exam will be scheduled by the Registrar. The final exam will cover concepts from the entire course and is in structure similar to the midterm exam. A grade of 50% or higher based on the weighted mean (25/35) of midterm and final is required to pass the course. The final has a similar format to the midterm.

For Graduates:

The 5 hr final exam will be scheduled by the Registrar. The final exam will cover concepts from the entire course and is in structure similar to the midterm exam – it will be partly overlapping with the undergraduate exam, but contain an assignment that requires more independent application of machine learning techniques to a practical problem. A grade of 50% or higher based on the weighted mean (25/35) of midterm and final is required to pass the course.

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 Academic Accommodation for Students with Disabilities policy can be found at: https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Accommodation_disabilities.pdf

Academic Consideration for Student Absence

Students will have up to two (2) opportunities during the regular academic year to use an on-line portal to self-report an absence during the semester, provided that the absence is no more than 48 hours in duration. Self-reported absences can be applied to a homework assignment. For this, the self-reported absence has to be reported to the online-portal before the due date of the homework. In this case, the student will receive a 48-hr extension for the submission of the homework – however, the homework still needs to be submitted.

If an absence is reported for the date of the midterm, the weight for the cumulative final will be increased to 60%. **There will be no make-up midterm.** Self-reported absences cannot be applied to the final, as the weight exceeds 30%.

If the conditions for a Self-Reported Absence are *not* met, students will need to provide a Student Medical Certificate if the absence is medical, or provide appropriate documentation if there are compassionate grounds for the absence in question. Students are encouraged to contact their Faculty academic counselling office to obtain more information about the relevant documentation.

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

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

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

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

You may also be eligible to write the Special Exam if you are in a “Multiple Exam Situation” (see http://www.registrar.uwo.ca/examinations/exam_schedule.html).

7. Academic Policies

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

In accordance with policy, https://wts.uwo.ca/identity/identities_and_access/activation.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.

On the midterm and exam, students are allowed to use any document and source on their computer and look up documents on the internet. *However they are not allowed to share documents, or communicate in any other way with other human or artificial intelligent agents during the exam.*

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>).

All of the remote learning sessions (Q&A) for this course will be recorded and made available on OWL. The data captured during these recordings may include your image, voice recordings, chat logs and personal identifiers (name displayed on the screen). The recordings will be used for educational purposes related to this course, including evaluations. The recordings may be disclosed to other individuals participating in the course for their private or group study purposes. Please contact the instructor if you have any concerns related to session recordings.

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
- All recorded sessions will remain within the course site or unlisted if streamed

Copyright Statement

Please be aware that all course materials created by the instructor(s) are copyrighted and cannot be **sold nor 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 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.

The policy on Accommodation for Students with Disabilities can be found here:

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic%20Accommodation_disabilities.pdf

The policy on Accommodation for Religious Holidays can be found here:

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

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>.

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/>.