Department of Mathematics & Statistics

Dsci 400

Machine Learning I

Winter 2022

Professor Information

Professor:	Bevan Ferreira	
Campus:	Kelowna	
Offices:	Online, via Zoom or similar	
E-mail:	bferreira@okanagan.bc.ca	
Office Hours:	Mon & Wed	15:30 - 16:30
	Tue & Thu	10:30 - 11:30
	Fri	11:30 - 12:30

Section Information

Section:	001		
Class Times:	Mon	09:30 - 11:00	Online
	Wed	09:30 - 11:00	Online
	Fri	10:00 - 11:00	Online

Calendar Description

DSCI 400-3-4
Machine Learning I
This course is a continuation of Data Wrangling and Visualization. Topics include: Exploratory graphs, plotting systems, hierarchical clustering, k-means clustering, dimension reduction, principle component analysis and singular valve decomposition. Students who have taken DSCI 101 for credit can not take DSCI 400 for further credit. (4,0,0)

Prerequisites:

- DSCI 300
- DSCI 310

Transfer Information

Please refer to the transfer guide, available online at http://www.bctransferguide.ca. Students are encouraged to save a copy of current transfer information for their own records.

Course Materials

The required texts for this course are:

Bradley Boehmke & Brandon Greenwell, **Hands-On Machine Learning with R**
CRC Press, 2020-02-01, ISBN# 978-1-138-49568-5
url: https://bradleyboehmke.github.io/HOML/

Roger D. Peng, **Exploratory Data Analysis with R**,
2020-05-01 https://leanpub.com/exdata

Note that additional materials in support of the key learning outcomes will be provided as needed, and may include additional class notes, audio-visual presentations, news articles, and research papers.

Course Content

The following is a synopsis of probable course content: We expect to cover material from Boehmke in Chapters 1-6, 17, 20 and 21. In addition, some or all of Peng, in Chapters 6, 7 & 9, will be reviewed to support a deeper understanding of appropriate and effective use of graphical analyses. We will study techniques for working with, and optimising, feature data, and use dimension reduction (for example, PCA/SVD, feature engineering) as well as a detailed analysis of some elementary clustering, classification and prediction techniques used in machine learning. These will include regression, k-means and hierarchical clustering. We will also perform analyses of model performance, and develop a working appreciation of the manifold risks associated with using models to reflect real-world phenomena.

Professionalism

Students must be aware that they should conduct themselves during the course **at all times** in accordance with the highest standards of professionalism and dedication to their craft. The modern workplace requires care and sensitivity to a variety of requirements that go beyond the mere production of calculated quantities. These requirements are usually represented in the working world in the form of performance evaluations, incident reports, and general assessments of a candidate's ongoing suitability for a position. Accordingly, students are strongly encouraged to steer well clear of any of the following:

- 'Snarky' or otherwise rude and belittling remarks in the class, or by email, either to your colleagues, or your professor
- Repeated requests for extensions on work due
- Poor attendance and/or participation
- Any and all appearances of plagiarism, or suspicions about your work. You must, at all times, give the appearance of the utmost trustworthiness and integrity in your work! It is your responsibility to be as transparent and forthcoming as possible, and to give confidence to any potential employer or client, that you are to be trusted with their sensitive and confidential data and analyses.

Learning Outcomes

The following are the anticipated learning outcomes of the course.

1. Show appropriate use of R plotting systems for appropriate exploratory analysis.
2. Create and appropriately interpret exploratory graphs.
3. Demonstrate appropriate use of techniques such as principle component analysis (PCA)/singular value decomposition (SVD).
4. Demonstrate appropriate use of k-means and hierarchical clustering techniques for various types of datasets.
5. Describe the core differences in analyses enabled by regression, classification, and clustering.
6. Produce appropriate interpretation of results from common data analyses.

Course Evaluation

Your grade in this course will be broken down as follows:

Professionalism & Participation	5%
Assignments/Quizzes	15%
Team Lectures	10%
Midterm mini-Project	30%
Final Project	40%
Total	100%

- **Professionalism & Participation** of students will be evaluated on a continual basis throughout the course. Repeated or particularly egregious breaches of the Professionalism requirements in this outline will result in a grade of '0' for this specific component.
- **Assignments** will be given to assist students in developing the skills necessary to successfully complete the term projects. Students are encouraged to collaborate, and consult one another on assignment work, in order to cultivate a collegial and professional working atmosphere. However your final presented work must be your own, compiled and delivered *independently*. Plagiarism (in particular, **cutting and pasting another's work, computations, writing, or code!**) into your own submission is a serious breach of the College Academic Integrity Policy, and will result in escalation to College administrative staff! *You have been warned!*. Late assignments, absent supporting documentation for compassionate or medical reasons, will receive a grade of '0'.
- **Quizzes** of shorter duration may be periodically assigned, where practical, time permitting. Quizzes are to be done alone, and unaided, within the time available. R and/or Excel may need to be used for quizzes.
- **Team Lectures** will take place during the semester. These will be lectures that you and your team jointly present to the class, on one of the topics in our machine learning course. This will greatly assist in acquiring the critical industry skill of quickly and efficiently researching and presenting new and complex information to non-technical stakeholders.
- **The Midterm mini-Project** will require that you work in groups or teams. This allows a formal team experience, and should prepare you partially for the demands of industry collaborative project work.
- **The Final Project** will require that you work in isolation, with NO collaboration or discussion with anyone (whether by email, phone, or in person, or any other form of personalised, responsive communication) other than your Professor. Failure to do so will constitute a serious breach of College Academic Integrity requirements, as well as the Professionalism requirements outlined previously. *For the purposes of all relevant College Policies, the Final Project is considered your "Final Exam"*.
  Note: The assessment of your Midterm and Final projects will take into account:
  - accuracy and effectiveness of methodology and analysis
  - applicability, layout, and suitability of appendices and code
  - accuracy and clarity of communication of presented results and conclusions
  - layout and presentation of your final report
  - effective and appropriate use of graphics

Department Policies

- All written assignments are due at the start of class. All assignments not handed in within the first five minutes of class will be assigned a grade of zero.
- The math department does not give make-up exams nor does the department allow students to write exams out of time without a valid medical or compassionate reason.
- It is expected the student attends all classes. If a student misses a class, it is the student's responsibility to get the material covered in class from their peers.
- No students may change sections of a course after the final add/drop date. If students wish to switch sections after the first day of class but before the final add/drop date, they should consult the chair of the Math department in order to not lose grade progress.
- Failure to achieve a grade of at least 45% on the final exam of a course will result in a failing grade for the course.
- When a student fails a course as a result of failing to achieve a final exam grade of 45%, the maximum grade that will be awarded is 49%.

- Calculators used for exams will satisfy the department's calculator policy. For Dsci 400, students are allowed a non-programmable, non-graphing scientific calculator. Graphing calculators are permitted at discretion of instructor.
- There will be no formula sheet given for this course.

Important Dates

Classes begin	Mon., Jan. 10
Last day to register	Fri., Jan. 21
Last day to receive a refund for course drop	Fri., Jan. 21
Last day to drop a course without a withdrawal being recorded on the student's record	Fri., Jan. 21
Statutory Holiday (no classes)	Mon., Feb. 21
Mid-semester study break (no classes)	Tue., Feb. 22 - Fri., Feb 25
Last day to withdraw without academic penalty	Fri., Mar. 11
Last day of classes	Thu., Apr. 14
Statutory Holiday (no classes)	Fri., Apr. 15 - Mon., Apr. 18
Final exam period	Tue., Apr. 19 - Fri., Apr. 29

Okanagan College Policies

Final Exam Policy: The procedures relating to final exams are significantly different than those that involve midterms. Final exam policy is determined by the college and a much more formal process is invoked should a student be unable to write the final exam. It is stated in the final exam policy that student travel plans are not a valid reason for writing an out-of-time final exam. As such, it is essential that you do not make travel plans prior to the final exam schedule being posted. The full final exam policy can be found at the following link.

http://webapps-5.okanagan.bc.ca/ok/Calendar/Examinations

The final exam schedule is determined by the Office of the Registrar and posted at the following link sometime around the middle of the semester.

https://www.okanagan.bc.ca/office-of-the-registrar/scheduling-office/scheduling-office#finalexam

Academic Integrity Policy: Okanagan College requires that all students are informed of the Academic Integrity Policy included in the College Calendar which can be found at the following link:

http://webapps-5.okanagan.bc.ca/ok/Calendar/AcademicIntegrity

College Student Conduct Policies: Okanagan College requires that students are informed of acceptable Student Conduct Policies included in the College Calendar which can be found at the following link:

http://webapps-5.okanagan.bc.ca/ok/Calendar/StudentConduct

Student Advising & Counselling

Accessibility Services collaborates with the academic departments of the college to arrange appropriate accomodation for students with a disability. If you require academic accomodation, please contact disability services. Contact, and other relevant information, can be found at:

https://www.okanagan.bc.ca/accessibility-services

Counselling Services has professionally trained staff that are available to assist students in coping with problem areas in their life (including: personal & career counselling, study skills) that interfere with maximizing their academic and social potential. For more information visit:

http://www.okanagan.bc.ca/counselling-services