DALHOUSIE UNIVERSITY
Inspiring Minds

Faculty of Science Course Syllabus
Department of Mathematics and Statistics
Introduction to data mining with R

STAT 2450 WINTER 2021

Instructor(s): Philippe Fullsack (email: Philippe.Fullsack@dal.ca)
Lectures: Brightspace
Laboratories: Brightspace

Course Description

This course provides an introduction to data mining and R programming, suited for science students. Data mining methods include a vast set of tools developed in different areas for identifying the patterns in data. Students will learn programming methods for manipulating and exploring data through learning the basic ideas of some clustering, regression and classification methods. No prior programming knowledge is assumed.

Course Prerequisites

MATH 1000.03 or MATH 1215.03 and either (STAT 1060.03 or MATH 1060.03) or (STAT 2060.03 or MATH 2060.03) or DISP

Course Objectives/Learning Outcomes

The broad goals of this course are twofold:
Firstly, to teach students R programming and some general scientific computing methods. Roughly the first half of course will be allocated to R programming, including: using R as a calculator, data types, data structures, external files, loops and flow control, conditional execution, user defined functions, and use of built in statistical/graphical functions.
Secondly, to introduce a number of concepts for statistical learning, including: multiple regression, CART, supervised vs unsupervised learning, the bias variance trade-off, performance evaluation, cross validation, and bootstrapping.

Course Materials

There is a Brightspace site for the course. This is where assignment information and announcements will be posted. The Brightspace site may also contains links to documents and additional activities placed on the web

server of the Department of Mathematics and Statistics.

Lecture videos will be posted as external learning resources on Brightspace. The Panopto software will be used to record/edit videos.
Students will be required to use statistical software as part of this course. The software used in the course is the state-of-the-art open-source statistical package R. R is available from www.r-project.org for Mac OS, Windows, and Linux. An online environment for R is also available at rstudio.cloud.

- Required Textbook: "Introduction to Statistical Learning with Applications in R"
  by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, published by Springer. Hard copy in bookstore, or available free online at:
  https://www-bcf.usc.edu/~gareth/ISL/index.html
- Assignments will require use of the statistical software package R, together with the program RStudio, and the Rmarkdown library.

Directions for installing the software are at:

https://mathstat.dal.ca/~fullsack/stat2450/Notes/Rintro.html

Textbook
An Introduction to Statistical Learning, with Applications in R, James, Witten, Hastie, Tibshirani, 2013, New York: Springer.

Course delivery

The course is divided in 7 modules. Teaching is asynchronous. All required material is posted online on Brightspace. Students are free to work on lectures, lab activities, assignments and quiz at their own pace as long within each module period (a week) as long as due dates of assignments and quizzes are respected.

Communication with students

The course uses discussion forums on Brightspace to allow students to post their questions. The course's TA is in charge of monitoring the questions. Discussion lists are organized by topics. Chats and forms may be used for additional interactions. The instructor will be available online for questions regarding lectures on Mondays (14h35-17h25).

Course Organization

The course is divided in 7 modules.
Each module includes 2 lectures, lab activities, one quiz and one assignment.
Modules will typically be delivered on a bi-weekly basis, except for Module 4 and 5 (one week each).

Component --	TOPIC --	Start	End
Module 1	R basics	Mon 11 Jan	Sun 24 Jan

Module 2	Functions/distributions	Mon 25 Jan	Sun 7 Feb
Module 3	t-confidence intervals	Mon 8 Feb	Sun 28 Feb
Module 4	Bootstrap	Mon 1 Mar	Sun 7 Mar
Module 5	Regression+Bootstrap	Mon 8 Mar	Sun 14 Mar
Module 6	Tree-based model	Mon 15 Mar	Sun 28 Mar
Module 7	Random forests	Mon 29 Mar	Thu 8 Apr

Course Content

Listed below in roughly chronological order are the topics to be covered. Note that these may be altered slightly as the term progresses.

- Presentation of R and Rstudio
- R as a calculator, data types, data frames, importing and exporting data
- Controlling the program flow in R
- User-defined functions in R
- Built-in functions for classical probability distributions
- Exploratory data analysis in R, plots and graphs
- Simulation of coverage of t-confidence interval
- Introduction to linear and polynomial regression
- Introduction to the bootstrap method
- Bootstrap confidence intervals for the slope of a linear regression
- Elementary mathematical analysis of the bootstrap
- Introduction to tree-based regression: CART model
- Model validation techniques: leave-one-out and cross-validation
- Introduction to the Random Forests model
- Case studies: examples of applications of models to real data sets.

Course Assessment

Component	Weight (% of final grade)	Date
Quizzes	30=6x5	7 quizzes, approximately bi-weekly
Assignments	70=7x10	7 assignments, approximately bi-weekly

Quizzes and assignments will be available on Brightspace. All dates and times refer to those displayed in Brightspace. Note that dates will be set to Halifax local time.
Students located in another time zone will have to use the time displayed in Brightspace, not their local civil time.
Assignments will be posted in R markdown format and students will be required to knit them to pdf in

Rstudio. They will be marked by the TA of this course, based on solutions provided by me.
I will eliminate the worst quiz on an individual basis (the 6 best quizzes out of 7 will be used for the final mark).

Quizzes will be marked automatically on Brightspace.

Each quiz is worth 5% of the final grade.
Each assignment is worth 10% of the final grade.

Component --	TOPIC --	Start	End
Quiz 1	R basics	Mon 11 Jan	Sun 24 Jan
Assignment 1	R basics	Mon 11 Jan	Sun 24 Jan
Quiz 2	Functions/distributions	Mon 25 Jan	Sun 7 Feb
Assignment 2	Functions/distributions	Mon 25 Jan	Sun 7 Feb
Quiz 3	t-confidence intervals	Mon 8 Feb	Sun 28 Feb
Assignment 3	t-confidence intervals	Mon 8 Feb	Sun 28 Feb
Quiz 4	Bootstrap	Mon 1 Mar	Sun 7 Mar
Assignment 4	Bootstrap	Mon 1 Mar	Sun 7 Mar
Quiz 5	Regression+Bootstrap	Mon 8 Mar	Sun 14 Mar
Assignment 5	Regression+Bootstrap	Mon 8 Mar	Sun 14 Mar
Quiz 6	Tree-based model	Mon 15 Mar	Sun 28 Mar
Assignment 6	Tree-based model	Mon 15 Mar	Sun 28 Mar
Quiz 7	Random forests	Mon 29 Mar	Thu 8 Apr
Assignment 7	Random forests	Mon 29 Mar	Thu 8 Apr

Other Course Requirements

Conversion of numerical grades to Final Letter Grades follows the Dalhousie Common Grade Scale

A+	(90–100)	**B+**	(77–79)	**C+**	(65–69)	**D**	(50–54)
A	(85–89)	**B**	(73–76)	**C**	(60–64)	**D**	< 50
A-	(80–84)	**B-**	(70–72)	**C-**	(55–59)	**D**	(50–54)

Course Policies

Credit cannot be given for late assignments.

ACCOMMODATION POLICY FOR STUDENTS

Students may request accommodation as a result of barriers related to disability, religious obligation, or any characteristic protected under Canadian Human Rights legislation. The full text of Dalhousie's Student Accommodation Policy can be accessed here:

http://www.dal.ca/dept/university_secretariat/policies/academic/student-accommodation-policy-wef-sep--1--2014.html

Students who require accommodation for classroom participation or the writing of tests and exams should make their request to the Advising and Access Services Centre (AASC) prior to or at the outset of the regular academic year. More information and the Request for Accommodation form are available atF www.dal.ca/access

ACADEMIC INTEGRITY

Academic integrity, with its embodied values, is seen as a foundation of Dalhousie University. It is the responsibility of all students to be familiar with behaviours and practices associated with academic integrity. Instructors are required to forward any suspected cases of plagiarism or other forms of academic cheating to the Academic Integrity Officer for their Faculty. The Academic Integrity website (http://academicintegrity.dal.ca) provides students and faculty with information on plagiarism and other forms of academic dishonesty, and has resources to help students succeed honestly. The full text of Dalhousie's Policy on Intellectual Honesty and Faculty Discipline Procedures is available here:

http://www.dal.ca/dept/university_secretariat/academic-integrity/academic-policies.html

STUDENT CODE OF CONDUCT

Dalhousie University has a student code of conduct, and it is expected that students will adhere to the code during their participation in lectures and other activities associated with this course. In general: "The University treats students as adults free to organize their own personal lives, behaviour and associations subject only to the law, and to University regulations that are necessary to protect

- the integrity and proper functioning of the academic and non–academic programs and activities of the University or its faculties, schools or departments;
- the peaceful and safe enjoyment of University facilities by other members of the University and the public;
- the freedom of members of the University to participate reasonably in the programs of the University and in activities on the University's premises;
- the property of the University or its members."

The full text of the code can be found here:

http://www.dal.ca/dept/university_secretariat/policies/student-life/code-of-student-conduct.html

SERVICES AVAILABLE TO STUDENTS

The following campus services are available to help students develop skills in library research, scientific writing, and effective study habits. The services are available to all Dalhousie students and, unless noted otherwise, are free.

Service	Support Provided	Location	Contact
General Academic Advising	Help with - understanding degree requirements and academic regulations - choosing your major - achieving your educational or career goals - dealing with academic or other difficulties	**Killam Library Ground floor** Rm G28 **Bissett Centre for Academic Success**	In person: Killam Library Rm G28 By appointment: - e-mail: advising@dal.ca - Phone: (902) 494-3077 - Book online through MyDal
Dalhousie Libraries	Help to find books and articles for assignments Help with citing sources in the text of your paper and preparation of bibliography	**Killam Library Ground floor** Librarian offices	In person: Service Point (Ground floor) By appointment: Identify your subject librarian (URL below) and contact by email or phone to arrange a time: http://dal.beta.libguides.com/sb.php?subject_id=34328
Studying for Success (SFS)	Help to develop essential study skills through small group workshops or one-on-one coaching sessions Match to a tutor for help in course-specific content (for a reasonable fee)	**Killam Library 3rd floor** Coordinator Rm 3104 Study Coaches Rm 3103	To make an appointment: - Visit main office (Killam Library main floor, Rm G28) - Call (902) 494-3077 - email Coordinator at: sfs@dal.ca or - Simply drop in to see us during posted office hours All information can be found on our website: www.dal.ca/sfs
Writing Centre	Meet with coach/tutor to discuss writing assignments (e.g., lab report, research paper, thesis, poster) - Learn to integrate source material into your own work appropriately - Learn about disciplinary writing from a peer or staff member in your field	**Killam Library Ground floor** Learning Commons & Rm G25	To make an appointment: - Visit the Centre (Rm G25) and book an appointment - Call (902) 494-1963 - email writingcentre@dal.ca - Book online through MyDal We are open six days a week See our website: writingcentre.dal.ca