DALHOUSIE UNIVERSITY
Inspiring Minds

Faculty of Science Course Syllabus
Department of Mathematics and Statistics
Data mining with R: a second course
STAT 3450 WINTER 2021

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

Course Description

This course teaches students how to use the R software environment to analyze data of various types. Students will be given the opportunity to write their own programs to explore data and solve meaningful problems. Emphasis will be given to a good understanding of supervised/unsupervised learning models, their founding principles, and to data processing and interpretation.

Course Prerequisites

STAT 2450

Course Objectives/Learning Outcomes

Module 1 Classification

Classification of data and discriminative learning methods

Introduction to Logistic regression, Linear and Quadratic Discriminant Analysis, Generalized linear models and K-nearest-neighbors' classifiers. Fitting and using these models in R.

Bias-variance trade-off. Occam's razor. Cross-validation and tuning of models hyperparameters.

Module 2 Regression

Variable selection and regularized linear regression methods

Best variables subset selection, shrinkage (LASSO, ridge-regression), principal components-regression, partial least-squares regression. Fitting and using these models in R.

Non-linear regression models: splines regression, generalized additive models

Module 3 Bagging

Tree-based methods. Short review of CART and Random Forests. Bootstrap aggregation.

Module 4 Boosting

Boosting-based supervised learning methods

Introduction to boosting of weak learners for regression or classification. Adaboost and gradient-boosting. Fitting and using these models in R.

Module 5 Dimension Reduction / Clustering

Principal component analysis, dimension reduction, manifold learning, building useful features for modelling, data augmentation. Fitting and using these models in R.

Quantitative representations of images, sounds, text, time-series, biomarkers.

Similarity and dissimilarity measures, multidimensional scaling.

Clustering methods (K-means, K-medioid).

Module 6 Support Vector Machines

Maximum margin classifiers.

Support Vector Machines. Kernelized classifiers.

Module 7 Neural networks

Introduction to neural networks. Fitting and using these models in R.

Course Materials

Textbook:

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

Additional References

- Data Mining with R: Learning with case studies, Luis Torgo, 2016, CRC Press

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 quizzes at their own pace within each module period, 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. Discussion lists are organized by topics. Instructor will monitor Discussion lists and be available online on Tuesdays and Thursdays, 11h35-12h55.

Course Organization

The course is divided in 7 modules.
Each module includes 2 lectures, lab activities, one quiz and one assignment.
Modules will be delivered on a biweekly basis (except Modules 3 and 4, weekly).

Component --	TOPIC --	Start	End
Module 1	Classification	Thu 7 Jan	Wed 20 Jan
Module 2	Regression	Thu 21 Jan	Wed 3 Feb
Module 3	Bagging	Thu 4 Feb	Wed 10 Feb
Module 4	Boosting	Thu 11 Feb	Wed 24 Feb
Module 5	Dimension Reduction / Clustering	Thu 25 Feb	Wed 10 Mar
Module 6	Support Vector Machines	Thu 11 Mar	Wed 24 Mar
Module 7	Neural networks	Thu 25 Mar	Wed 7 Apr

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.
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	Classification	Thu 7 Jan	Wed 20 Jan
Assignment 1	Classification	Thu 7 Jan	Wed 20 Jan
Quiz 2	Regression	Thu 21 Jan	Wed 3 Feb
Assignment 2	Regression	Thu 21 Jan	Wed 3 Feb
Quiz 3	Bagging	Thu 4 Feb	Wed 10 Feb
Assignment 3	Bagging	Thu 4 Feb	Wed 10 Feb
Quiz 4	Boosting	Thu 11 Feb	Wed 24 Feb
Assignment 4	Boosting	Thu 11 Feb	Wed 24 Feb
Quiz 5	Dimension Reduction/Clustering	Thu 25 Feb	Wed 10 Mar
Assignment 5	Dimension Reduction/Clustering	Thu 25 Feb	Wed 10 Mar
Quiz 6	Support Vector Machines	Thu 11 Mar	Wed 24 Mar
Assignment 6	Support Vector Machines	Thu 11 Mar	Wed 24 Mar
Quiz 7	Neural networks	Thu 25 Mar	Wed 7 Apr
Assignment 7	Neural networks	Thu 25 Mar	Wed 7 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