DALHOUSIE UNIVERSITY

Faculty of Science Course Syllabus
Department of Mathematics and Statistics
Introduction to Data Mining with R – STAT 2450
Winter 2018

Instructor(s): *Chaoyue Liu Chaoyue.Liu@dal.ca*

Lectures: *MW 1:05 pm-2:25 pm SIR JAMES DUNN 135*

Laboratories: *None*

Tutorials: *None*

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 and either STAT/MATH 1060 or STAT/MATH 2060

Course Objectives/Learning Outcomes

The main goal of this course is to teach students R programming and general scientific computing methods. The more specific topics are following:

- *Explain the key differences between the tasks of classification, clustering, regression, and dimensionality reduction*
- *Identify the key differences between supervised and unsupervised learning paradigms*
- *Estimate the effects of hyperparameters on the resulting performance of data mining methods*
- *Understand the model complexity with regards to the bias-variance trade-off*
- *Understand the concepts over-fitting and under-fitting*
- *Recognize how to evaluate the performance of predictive models*
- *Estimate the effects of hyperparameters on the resulting performance of data mining methods*

- *Propose a suitable visualization design for a particular combination of data characteristics and application tasks*
- *Design data mining experiments using R and existing data mining tools*
- *Apply file-operations on given data sets for reading and writing with R*
- *Manipulate and interpret the data frame in R*
- *Explain and use the concept of loops to perform repetitive tasks*
- *Write reasonably-complex R scripts to solve common data tasks*
- *Apply the Nearest Neighbours method for supervised learning tasks*
- *Apply the CART-based decision tree learning method for supervised learning tasks*
- *Apply the K-fold cross-validation and hold-out validation techniques for assessing the performance of a predictive model*
- *Apply the grid search method for hyperparameter optimization*
- *Understand how support vector machines discover an optimal hyperplane for classification based tasks*
- *Apply the random forest for analyzing feature importance*
- *Apply the single-layer perceptron learning algorithm for constructing a classifier*
- *Understand the backpropagation algorithm for training the weights of a feed-forward neural network*
- *Apply the K-means algorithm for discovering centroid-based clusters*
- *Apply principal component analysis to project data onto lower dimensions*

Required Course Materials

Textbook: "Introduction to Statistical Learning with Applications in R"
by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani
published by Springer (Available freely online)

DALHOUSIE UNIVERSITY

Course Assessment

Component	Weight (% of final grade)	Date
Assignments	*50*	*4 – 5 assignments, TBA*
Midterm exam	*20*	*In class, TBA*
Final exam	*30*	*TBA*

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)	**F** (<50)
A- (80-84)	**B-** (70-72)	**C-** (55-59)	

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. **Student Accommodation Policy:** http://www.dal.ca/campus_life/student_services/academic-support/accessibility.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 at 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.

Policy on Intellectual Honesty and ***Faculty Discipline Process***:
https://www.dal.ca/dept/university_secretariat/academic-integrity.html

DALHOUSIE UNIVERSITY

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

COPYRIGHT

All members of the Dalhousie community are expected to comply with their obligations under Canadian copyright law. Dalhousie copyright policies and guidelines, including our Fair Dealing Guidelines, are available at http://www.dal.ca/dept/copyrightoffice.html.

SERVICES AVAILABLE TO STUDENTS

The following campus services are available to all Dalhousie students. Unless noted otherwise, the services 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 - e-mail Coordinator at: sfs@dal.ca or - 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 a tutor to discuss writing assignments (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 Writing Centre in the Killam Learning Commons (Rm G40) and book an appointment - Call (902) 494-1963 - e-mail writingcentre@dal.ca - Book online through MyDal We are open six days a week **See our website: writingcentre.dal.ca**