Carleton University
Department of Systems and Computer Engineering
SYSC 5405 / BIOM 5405 Pattern Classification and Experimental Design
Fall 2018
Course Information

Instructors:
Dr. A. Boyle, boyle@sce.carleton.ca, CB6105A
Office Hours: By appointment with additional hours posted on cuLearn

Course Objectives:
Introduction to a variety of supervised and unsupervised pattern classification techniques with emphasis on correct application. Statistically rigorous experimental design and reporting of performance results. Case studies will be drawn from various fields including biomedical informatics.

Prerequisites: Undergraduate introductory probability and statistics. (Plus graduate standing, permission of the instructor or the department.)

Course Website:
cuLearn will be used for the course webpage, which is accessible at culearn.carleton.ca. University of Ottawa students need to complete a form and submit it to the FGPA to gain access to CULearn for the term.

Students are required to check CULearn often for course updates. Supplementary lecture notes will be posted there for student use. Note that reading the supplementary lecture notes only is NOT ENOUGH to pass this course! The single best predictor of student performance is attendance at lectures.

Primary Textbook:

Related Textbooks (not mandatory):
- Weiss, Kulikowski, Computer Systems That Learn, Morgan Kaufmann, 1991
- James, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer 2013
  - Current literature describing applications of pattern classification
  - Freely available MATLAB/JAVA/C++ implementations of pattern classification methods (e.g. Netlab and WEKA)
  - Additional materials may be made available on the course website

*Items 1, 2, and 3 are on reserve at the CU Library.*

Grading:
- Problem Assignments: 20%
- Literature Review(s): 20%
- Term Project: 30%
- Final Exam: 30%

Note that only portions of the assignment(s) may be used for purposes of evaluation.
Important Notes:

Expectations for this course are common to most university courses:

You are expected to attend all lectures: If a student is absent from a lecture, it is up to the student to obtain missed lecture material from colleagues in the course.

Students who miss a deadline: due to illness, injury or other extraordinary circumstances are responsible for immediately informing the instructor concerned and for making alternate arrangements with the instructor and in all cases this must occur no later than three (3.0) working days after the term work was due. The alternate arrangement must be made before the last day of classes in the term as published in the academic schedule. For more information, see the Academic Regulations of the University, Section 2.6, Deferred Term Work.

Students who miss a deadline without a very good reason may be denied the right to submit their work or may be penalized (e.g. 10% per day).

Academic Accommodation: You may need special arrangements to meet your academic obligations during the term because of disability, pregnancy or religious obligations. Please review this course outline promptly and write to me with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist.

The Paul Menton Centre for Students with Disabilities (PMC) provides services to students with Learning Disabilities (LD), psychiatric/mental health disabilities, Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD), chronic medical conditions, and impairments in mobility, hearing, and vision. If you have a disability requiring academic accommodations in this course, please contact PMC at 613-520-6608 or pmc@carleton.ca for a formal evaluation. If you are already registered with the PMC, contact your PMC coordinator to send me your Letter of Accommodation at the beginning of the term, and no later than two weeks before the first in-class scheduled test or exam requiring accommodation (if applicable). Requests made within two weeks will be reviewed on a case-by-case basis. After requesting accommodation from PMC, meet with me to ensure accommodation arrangements are made. Please consult the PMC website (http://www.carleton.ca/pmc) for the deadline to request accommodations for the formally-scheduled exam (if applicable): https://carleton.ca/pmc/students/dates-and-deadlines/

Academic Integrity: Plagiarism and cheating at the graduate level are viewed as being particularly serious and the sanctions imposed are accordingly severe. Students are expected to familiarize themselves with and follow the Carleton University Student Academic Integrity Policy (See https://carleton.ca/registrar/academic-integrity/). The Policy is strictly enforced and is binding on all students. Plagiarism and cheating—presenting another’s ideas, arguments, words or images as your own, using unauthorized material, misrepresentation, fabricating or misrepresenting research data, unauthorized co-operation or collaboration or completing work for another student— weaken the quality of the graduate degree. Academic dishonesty in any form will not be tolerated. Students who infringe the Policy may be subject to one of several penalties including: expulsion; suspension from all studies at Carleton; suspension from full-time studies; and/or a reprimand; a refusal of permission to continue or to register in a specific degree program; academic probation; or a grade of Failure in the course.

Health and Safety: Every student should have a copy of our Health and Safety Manual. An electronic version of the manual can be found at http://www.sce.carleton.ca/courses/health-and-safety.pdf
Tentative Week-by-Week Schedule:

Weeks 1-2 Introduction to Pattern Classification. Data pre-processing, analysis, outlier detection, and transformations. Experiment design (feature selection and dimensionality, selecting classifier structure, test protocols, cross-validation, data partitioning, etc.). Introduction to hypothesis testing. Avoiding fundamental errors of testing on the training set and training on the test set.

Week 3 Reporting results. How to accurately and honestly report classification system performance. True error vs. apparent error. Confidence intervals, statistical tests to compare methods, receiver operator characteristic curves, sensitivity, specificity, confusion matrices, P-values. Class imbalance. Critical assessment of reported results in the literature.

Note that by introducing experiment design and reporting of results early in the course, applications discussed during all subsequent topics will be evaluated using these fundamental principles. The remainder of the course will survey a number of approaches to pattern classification. The depth of coverage will vary and will depend on time available. Relevant applications of pattern classification techniques from the literature will be discussed.


Week 5 Maximum likelihood and Bayesian parameter estimation. Non-parametric techniques such as Parzen windows, probabilistic neural networks, and K-nearest neighbour estimators and classifiers.

Week 6 Decision trees and decision forests. Training, pruning, splitting, and stopping criteria.

Week 7 Linear and nonlinear discriminant analysis. Linear discriminant functions and decision surfaces. Perceptron criterion, relaxation procedures, and MSE procedures. Generalized linear discriminant functions and support vector machines (briefly).

Week 8 Neural networks: network structure, feedforward operation and classification, backpropagation training.

Week 9 Markov chains, hidden Markov models, and expectation maximization.

Week 10 Meta-learner and re-sampling approaches including bagging and boosting. Combination of multiple experts: voting strategies and cascaded classifiers. Learning with queries.

Week 11 Unsupervised clustering (hierarchical, K-means, SOMs). Mixture densities, criterion functions for clustering, the number of cluster problem, and cluster validation.

Week 12 Student project presentations and competition. Review.