



Wentworth Institute of Technology

College of Engineering and Technology

COMP4050 – Machine Learning
Fall 2015

Instructor	Nate Derbinsky
Office	Dobbs 140 MTWF 4-5PM and by appointment
Contact	(617) 989-4287 derbinskyn@wit.edu http://derbinsky.info
Credits/Hours	3/2/4

COURSE DESCRIPTION:

Introduction to the field of machine learning. This course focuses on algorithms to help identify patterns in data and predict or generalize rules from these patterns. Topics include supervised learning (parametric/non-parametric algorithms, kernels, support vector machines), model selection, and applications (such as speech and handwriting recognition, medical imaging, and drug discovery). Students who have basic programming skills and who have taken a course in probability are encouraged to take this course.

COURSE PREREQUISITES/COREQUISITES:

Prerequisites:

- MATH2100 Probability and Statistics for Engineers
- COMP1000 Computer Science I

REQUIRED TEXTBOOK(S):

- Harrington, Peter. *Machine Learning in Action*, 1st ed. Manning, 2012 (ISBN-13: 978-1617290183)
- Bishop, Christopher M.. *Pattern Recognition and Machine Learning*, 1st ed. Springer, 2007 (ISBN-13: 978-0387310732)

THE COLLEGE BOOKSTORE:

Location: 103 Ward Street Boston MA 02115
Telephone: (617) 445-8814

RECOMMENDED LEARNING MATERIALS:

- Sutton, Richard S., Barto, Andrew G.. *Reinforcement Learning: An Introduction*, 1st ed. A Bradford Book, 1998 (ISBN-13: 978-0262193986¹)
- Wu, Xindong, Kumar, Vipin. *The Top Ten Algorithms in Data Mining*, 1st ed. Chapman and Hall/CRC, 2009 (ISBN-13: 978-1420089646²)
- Russell, Stuart, Norvig, Peter. *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2009 (ISBN-13: 978-0136042594)
- Witten, Ian H., Frank, Eibe, Hall, Mark A.. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Morgan Kaufmann, 2011 (ISBN-13: 978-0123748560)
- Data Mining with Weka MOOC³
- Machine Learning, CMU, Tom Mitchell⁴
- Machine Learning, Stanford, Andrew Ng⁵

COURSE LEARNING OUTCOMES:

At the completion of this course, the student should be able to:

- I. Use computational techniques for data transformation
- II. Predict rules from patterns in data
- III. Explain and use important vector operations
- IV. Apply machine learning algorithms to real datasets and evaluate their performance
- V. Explain bias vs. variance trade-off
- VI. Explain what training data is vs. testing data and how they are used to learn
- VII. Evaluate a learning cycle by using multiple training and test sets in resampling schemes such as bootstrapping and cross-validation

¹<http://webdocs.cs.ualberta.ca/~sutton/book/ebook/the-book.html>

²<http://www.cs.umd.edu/~samir/498/10Algorithms-08.pdf>

³<http://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/>

⁴http://www.cs.cmu.edu/~tom/10701_sp11/

⁵https://www.youtube.com/view_play_list?p=A89DCFA6ADACE599

INSTRUCTIONAL METHODOLOGIES:

This course will be problem based and interactive. Students are expected to read the textbook, and participate by asking and responding to questions during class. There will be frequent homework assignments, as well as regular quizzes. For individual attention, students are encouraged to attend office hours. This syllabus and other relevant course handouts will be posted on Blackboard (<http://bb.wit.edu>).

ATTENDANCE POLICY:

Your attendance is expected at every class. Please arrive on time to every class: attendance will be taken at the beginning of class and late arrivals will be recorded as absences. If you have a legitimate reason for missing a class, send the instructor an email, preferably ahead of time, in order to be excused for that class. If you do have to miss a class, then it is your responsibility to learn the material covered and to check on any announcements that were made.

Students are expected to attend classes regularly, take tests, and submit papers and other work at the times specified by the instructor. Students who are absent repeatedly from class or studio will be evaluated by faculty responsible for the course to ascertain their ability to achieve the course objectives and to continue in the course. Instructors may include, as part of the semester's grades, marks for the quality and quantity of the student's participation in class. At the discretion of the instructor, a student who misses 15 percent of class may be withdrawn from the course by the instructor. A grade of **WA** will appear on the student's official transcript as a result.

GRADING POLICY:

Homework	40%
Quizzes	40%
Final Project	20%

Homework will be posted and submitted via Blackboard. You will turn in a combination of source code and worked-out problems (preferably \LaTeX), and you will typically have about 2 weeks to work on multiple problems. The intent is for you to gain experience working with data and evaluating the performance of machine learning techniques.

Homework 0: Mandatory! Schedule (via e-mail) and attend a 5-minute, one-on-one appointment with the instructor by the end of the second week of class.

Quizzes will be given regularly in class, typically once per week. Unless otherwise specified, quizzes will be closed-book, closed-notes. The intent is to make sure you keep up with the reading, know the vocabulary, understand applicability of the methods, and grasp the concepts of lectures/labs. There will be no midterm or final exam.

Final Project components (see the specification document) will be submitted via Blackboard. The intent is for you to get in-depth experience with a dataset, an algorithm, a paper, and/or the theory/application of machine learning.

WENTWORTH GRADING SYSTEM:

Grade	Definition	Weight	Numerical
A	Student learning and accomplishment far exceeds published objectives for the course/test/assignment and student work is distinguished consistently by its high level of competency and/or innovation.	4.00	96 – 100
A-		3.67	92 – 95
B+	Student learning and accomplishment goes beyond what is expected in the published objectives for the course/test/assignment and student work is frequently characterized by its special depth of understanding, development, and/or innovative experimentation.	3.33	88 – 91
B		3.00	84 – 87
B-	Student learning and accomplishment meets all published objectives for the course/test/assignment and the student work demonstrates the expected level of understanding, and application of concepts introduced.	2.67	80 – 83
C+		2.33	76 – 79
C		2.00	72 – 75
C-	Student learning and accomplishment based on the published objectives for the course/test/assignment were met with minimum passing achievement.	1.67	68 – 71
D+		1.33	64 – 67
D		1.00	60 – 63
F	Student learning and accomplishment based on the published objectives for the course/test/assignment were not sufficiently addressed nor met.	0.00	< 60

ADD/DROP:

Students should check the academic calendar to confirm the add/drop deadline. Dropping and/or adding courses is done online. Courses dropped in this period are removed from the student's record.

Non-attendance does not constitute dropping a course. If a student has registered for a course and subsequently withdraws or receives a failing grade in its prerequisite, then the student must drop that course. In some cases, the student will be dropped from that course by the Registrar. However, it is the student's responsibility to make sure that he or she meets the course prerequisites and to drop a course if the student has not successfully completed the prerequisite. The student must see his or her academic advisor or academic department chair for schedule revision and to discuss the impact of the failed or withdrawn course on the student's degree status.

MAKE-UP POLICY:

All assignments have a specific due date and time. Submissions will be accepted *up to one day* after the deadline with a 50% penalty. The assignment will be graded and returned as normal, but the grade will be recorded as half of what was earned. For example, an on-time submission might receive a grade of 90 points. The same assignment submitted after the deadline would receive 45 points (90×0.5).

Students who miss scheduled exams will not, as a matter of course, be able to make up those exams. If there is a legitimate reason why a student will not be able to complete an assignment on time or not be present for an exam, then they should contact the instructor beforehand. Under extreme circumstances, as decided on a case-by-case basis by the instructor, students may be allowed to make up assignments or exams without first informing the instructor.

ACADEMIC SUPPORT:

The Learning Center (TLC) assists all Wentworth students in the areas of math, science, technical courses specific to majors, and writing. In this student-based learning environment, students can receive individual help with their studies, meet and work in study groups, attend workshops on a wide variety of subjects and find resources to assist them in meeting their goals for academic success. It includes tutors in many subjects, writing assistance and workshops focused on helping good students become great students. Make appointments at <http://www.wit.edu/tlc> or through LConnect.

Applied Math Facilitated Study Groups (FSGs) will be held one day per week, Tuesdays from 3:00 to 6:00 pm, in the Ira Allen Forum starting 9/8/2015 and ending on 12/1/2015. This is a place where you can: (1) ask questions and get help in your math classes, (2) work in groups, or (3) get challenged. At least one Applied Math professor, along with two student teaching assistants, will be there to help you. This is for all levels of mathematics (including those related to Machine Learning). No appointment is necessary – just drop in anytime.

ACADEMIC HONESTY STATEMENT:

“Students at Wentworth are expected to be honest and forthright in their academic endeavors. Academic dishonesty includes cheating, inventing false information or citations, plagiarism, tampering with computers, destroying other people’s studio property, or academic misconduct” (Academic Catalog). See your catalogue for a full explanation.

STUDENT ACCOUNTABILITY STATEMENT:

Behavior unbecoming a student is any violation of a published Wentworth policy in an academic environment, and/or any behavior that individual faculty or staff determines is unacceptable in his or her classroom, laboratory, or other academic area or function. Behavior unbecoming a student in an academic environment will not be tolerated. Violations of behavioral expectations may be forwarded to the Office of Community Standards for disciplinary action.

Wentworth takes violations of academic dishonesty and misconduct very seriously. Sanctions for such violations include, but are not limited to, a grade of “F”, removal from a course, Institute suspension, or Institute expulsion.

DISABILITY SERVICES STATEMENT:

Any student who thinks s/he may require a disability-related accommodation for this course should contact Disability Services privately to discuss their specific needs. Disability Services coordinates reasonable accommodations for students with documented disabilities. They are located in Watson Hall 003 (the Center for Wellness and Disability Services) and can be contacted at 617-989-4390 or counseling@wit.edu. For more information on acceptable documentation and the Disability Services process, visit the Disability Services website at <http://www.wit.edu/disabilityservices>.

COLLEGE OF THE FENWAY STUDENTS:

If you are enrolled in this course through COF Cross Registration, notify your course instructor. Please provide her/him with your email address to be sure that you receive course information in a timely way. You should also discuss how to access online applications that might be used in the course.

WEEKLY SCHEDULE:

The following schedule is tentative and subject to change (including topics, assignments, and quizzes). It will benefit you greatly to complete the assigned reading *before* attending the lecture.

Readings reference the following texts:

- **MLiA:** Machine Learning in Action
- **PRML:** Pattern Recognition and Machine Learning
- **RL:** Reinforcement Learning: An Introduction
- **TT:** The Top Ten Algorithms in Data Mining
- **AMA:** Artificial Intelligence: A Modern Approach (optional)

Week	Topic	Reading	Assignments/Notes
1	Introduction to Machine Learning	MLiA: 1 PRML: 1 ⁶ AMA: 18.1, 18.2	
2	k NN, Evaluation	MLiA: 2 PRML: 2.5 TT: 8 AMA: 18.4, 18.8	Quiz 1, HW1 Due
3	Decision Trees	MLiA: 3 PRML: 14.4 TT: 1 AMA: 18.3	Quiz 2

⁶1.0, 1.4. As needed for review: 1.2-1.2.4

Week	Topic	Reading	Assignments/Notes
4	Naïve Bayes	MLiA: 4 PRML: 8.1, 8.2 TT: 9 AMA: 13, 20.2.2	Quiz 3, HW2 Due
5	Logistic Regression	MLiA: 5 PRML: 4.3 AMA: 18.6	Quiz 4 PRJ: Interest Statement
6	Linear Regression	MLiA: 8 PRML: 3.1, 3.2 AMA: 18.6	Quiz 5, HW3 Due
7	k -Means Clustering	MLiA: 10 PRML: 9.1 TT: 2	Quiz 6
8	Reinforcement Learning	PRML: 1.0 RL: 1 – 6 AMA: 21	Quiz 7, HW4 Due
9	Apriori, AdaBoost	MLiA: 7, 11 PRML: 14.2, 14.3 TT: 4, 7 AMA: 18.10	Quiz 8 PRJ: Proposal
10	Artificial Neural Networks	PRML: 5 AMA: 18.7	Quiz 9, HW5 Due
11	Dimensionality Reduction	MLiA: 13, 14 PRML: 12.1	Quiz 10 PRJ: Update
12	Support Vector Machines	MLiA: 6 PRML: 7.1 TT: 3 AMA: 18.9	Quiz 11, HW6 Due
13	Advanced Topics ⁷		Quiz 12
14	Project Presentations		Final Project Due

⁷Example topics: EM, graphical models, scaling, ...