COM-3920-341: Machine Learning

Course Information	
Prerequisite	COM 3900 Mathematics for Machine Learning
Description	This course takes an application driven approach to current topics in machine learning. The course covers supervised learning, unsupervised learning, semi-supervised learning, and several other learning settings. We will cover popular algorithms and will focus on how statistical learning algorithms are applied to real world applications. Students will implement several learning algorithms throughout the semester. The goal of this course is to provide students with the basic tools they need to approach various applications, such as Biology/Bioinformatics, Information Retrieval, Natural Language Processing, Speech Processing, and Vision. We will focus on fundamental methods applicable to all applications.
Course Outcomes	• Students will learn the fundamentals of machine learning • Students will learn to implement machine learning algorithms • Students will learn to evaluate how to apply machine learning to different settings

Major Topics Covered in Course	• Overview of Machine Learning (ML) applications
	• Linear Regression background
	• Supervised Learning (up to midterm exam)
	  ○ Logistic Regression, introducing various classification methods
	  ○ Perceptron and related online learning methods
	  ○ Support Vector Machines: max-margin classification and optimization
	  ○ Kernel Methods, especially dual optimization
	  ○ Decision Trees: construction, pruning, and over-fitting
	  ○ Boosting, especially ensemble methods
	  ○ Deep Learning
	• Unsupervised Learning (after midterm exam)
	  ○ Clustering
	  ○ Expectation Maximization techniques
	  ○ Dimensionality Reduction and Principal Components Analysis
	  ○ Graphical Methods
	  ○ Structured Prediction
	• Practical Application of Machine Learning

Text Book(s)	**Required**: Kevin Murphy, *Machine Learning, a Probabilistic Perspective*, 1st edition, MIT Press (2012) [author disavows the eBook version sold on Amazon, only buy an eBook from MIT Press] **Supplemental:** Christopher Bishop, *Pattern Recognition and Machine Learning,* Springer (2011 printing)
Assignments	Students will complete 4 major assignments applying established Machine Learning techniques in different problem domains and analyzing the resulting performance. These assignments will involve mathematical analysis, programming, and writing. Different techniques and data sets will be selected each term.
Assignment Grading	The assignments will be graded as follows: Correctness and readability of program code: 40% Mathematical analysis of results: 40% Clarity and cogency of writing style: 20% **Assignments submitted after the due date & time receive a grade of zero.**

Exams	There will be a midterm exam in the 8th week as well as a final exam.
Components of Student's Grade	Assignments: 50% Final exam: 30% Midterm exam: 20%
Grading Scale:	A = 93-100% B+ = 87-89% C+ = 77-79% D+ = 67-69% A- = 90-92% B = 83-86% C = 73-76% D = 65-66% B- = 80-82% C- = 70-72% F = 64 and lower
Credits	3
Weekly Schedule	Two sessions a week, 75 minutes each.
Attendance Policy	• Attendance of every session is mandatory. • Every unexcused absence incurs a penalty of one point off the final exam. • Every third unexcused absence additionally incurs the penalty of the student's final grade in the course being lowered by two letter "places" (e.g. from an A- to a B, from a B+ to a B-, etc.)
Y.C. C.S. Department Academic Integrity Policy	**If you need help with any aspect of any Y.C. C.S. course, please reach out to your professor and/or TA - we are there to help. Do not under any circumstances resort to cheating or plagiarizing in any way.** All academic integrity cases in Y.C. C.S. classes will be handled as follows: **1)** Every case will be referred to the dean's office for investigation and disciplinary measures - no exceptions **2)** The first time a student is caught cheating or plagiarizing on any part of any work item (e.g. a homework assignment, exam, etc.) for any Y.C. C.S. course, he will receive a zero on the work item on which he cheated or plagiarized and have his final grade in the course lowered by an entire letter (e.g. from B+ to C+.) Repeat offenders, whether they repeat in a single semester or across multiple semesters, will be dealt with more stringently. These penalties have been, and will be, applied even if it means a senior not graduating and/or a student having to take the course all over again. For more information, please see Yeshiva College's Academic Integrity Policy.

Students With Disabilities

The Office of Academic Support provides services and resources designed to help students on Wilf Campus develop more efficient and effective study skills and strategies. Individual support is available in areas such as time management and organization, active reading, note-taking, exam preparation and test-taking skills. The office is located in Furst Hall, Suite 412. To schedule an appointment, call 646-592-4285 or email academicsupport.wilf@yu.edu.

Weekly Schedule	
Week	**Topics**
1	Overview of Machine Learning: history, relation to classical statistics **Text:** Chapter 1;

2	**Linear Regression:** model specification issues, maximum likelihood estimation, ridge regression **Text:** Chapter 7, except 7.4 and 7.6
3	**Logistic Regression:** model fitting methods (steepest descent, Newton's, L1 and L2 regularization, etc.) **Online learning algorithms**: structured prediction, regret minimization, stochastic optimization, risk minimization, LMS algorithm, perceptron algorithm, relation to Bayesian techniques. **Text:** Chapter 8.1-8.3.6, 8.5, 13.3,
4	**Support Vector Machines:** uses in regression and classification, optimization, choice of parameter C, probabilistic interpretation **Kernel Methods:** RBF kernels, Mercer kernels, Matern kernels, Linear kernels, String kernels, Kernels derived from probabilistic generative models; the "kernel trick" and its uses: nearest neighbor classification, K-medoids clustering, ridge regression, and PCA **Text:** Chapter 14.1, 14.2, 14.5, supplemental notes by Ng
5	**Decision Trees:** basis in information theory, when appropriate to use, growing and pruning, detecting and avoiding over-fitting; pros and cons with respect to other ML techniques; random forests; ID3 and related algorithms; use of bias; training with incomplete data **Boosting:** why it works so well, specific types of boosting, ensemble learning **Text:** Chapters 2.8, 16.2, 16.4, 16.6 , supplemental notes by Mitchell, Yoav & Schapire
6	**Deep Learning:** background in feed-forward neural networks (multilayer perceptrons) and Restricted Boltzmann machines; deep generative models (sigmoid networks, Boltzmann machines, belief networks) **Text:** Chapters 16.5, 27.7, 28.1-28.2
7	**Deep Learning (continued):** training of deep networks, applications of deep networks **Midterm Review** **Text:** Chapters 28.3-28.4
8	**Midterm Exam** **Clustering:** measures of dissimilarity; k-means clustering, clustering in mixture models **Text:** Chapters 25.1, 11.1-11.3
9	**Expectation Maximization:** basic idea, Jensen's inequality, the EM algorithm, EM applied to various distributions (mixture of Gaussians), theoretical basis **Text:** Chapters 11.4, supplemental notes by Ng
10	**Dimensionality Reduction:** classical Principal Components Analysis, PCA theorem; Singular Value Decomposition, Probabilistic PCA, EM algorithm for PCA, PCA for categorical data, PCA for paired and multi-view data **Text:** Chapters 12.2, 12.4, 12.5
11	**Graphical Techniques:** Bayesian nets, Directed Gaussian Models (DGM), plate notation, d-separation, Markov random fields, Hammersley-Clifford theorem and its effects, training maxent models, pseudo-likelihood, stochastic maximum likelihood learning **Text:** Chapter 10, 19.1-19.5

12	**Graphical Techniques (continued):** belief propagation (serial, parallel, Gaussian, and other variants), variable elimination, junction trees, intractability of exact inference, approximate inference methods, max sum and max product **Text:** Chapter 20
13	**Structured Prediction:** Markov and Hidden Markov (HMM) models, inference algorithms (forward, Viterbi, backward filtering, etc.), training methods, Baum-Welch algorithm, model selection, Conditional Random Fields (CRF), label-bias problem, uses in handwriting recognition, protein analysis, and stereopsis **Text:** Chapters 17.1-17.5, 19.6-19.7, CRF tutorial paper by Sutton & McCallum