ONLINE MASTER OF SCIENCE IN ANALYTICS
ISYE/CSE 6740 – COMPUTATIONAL DATA ANALYSIS / MACHINE LEARNING I
TENTATIVE SYLLABUS (SUBJECT TO CHANGE), SUMMER 2020
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

PROFESSOR: Yao Xie; yao.xie@isye.gatech.edu
Professor Office Hour: Wed 9-9:30pm.

TEACHING ASSISTANTS:

(HEAD TA) Alfie Pengfei Chen, pfchen@gatech.edu

1. Pornpawee Bumpensanti, pornpawee@gatech.edu, Office Hour
2. Christopher Angelo Ginart, cginart3@gatech.edu, Office Hour
3. Jennifer Ann Glenski, jglenski@gatech.edu, Office Hour
4. Felipe Lopez, flopez32@gatech.edu, Office Hour
5. Anthony Kyobin Song, asong45@gatech.edu, Office Hour
6. Zhongmin Tao, ztao48@gatech.edu, Office Hour
7. Jianyong Wu, jwu610@gatech.edu, Office Hour

PREREQUISITES

- Undergraduate level probability, linear algebra, and statistics.
- MATLAB or Python.

COURSE DESCRIPTION

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. The course is designed to answer the most fundamental questions about machine learning: What are the most important methods to know about, and why? How can we answer the questions such as “is this method better than that one' using asymptotic theory”? How can we answer the question 'is this method better than that one' for a specific dataset of interest? What can we say about the errors our method will make on future data? What's the “right” objective function? What does it mean to be statistically rigorous?

This course is designed to give graduate students a thorough grounding in the methods, theory, mathematics and algorithms needed to do research and applications in machine learning. The course covers topics from machine learning, classical statistics, and data mining. Students entering the class with a pre-existing working knowledge of probability, statistics and algorithms will be at an advantage, but the class has been designed so that anyone with a strong numerate background can catch up and fully participate. Some experience with coding is expected (MATLAB or Python.)

For detailed course topics, please see the tentative course schedule.

LEARNING OBJECTIVES

After taking this course, students should be able to:

- Gain thorough understanding in the methods, theory, mathematics and algorithms needed to do research and applications in machine learning.
- Implementing and use machine learning algorithms.
- Gain experience with analyzing real data.

TEXTBOOKS/READINGS

- **Textbook**: The course material will be based on lectures slides provided in the course.
- **Recommended References**:
  - (PRML) Pattern recognition and Machine Learning, Christopher M. Bishop.
  - (ESL) The elements of Statistical Learning: Data Mining, Inference, and Predictions, 2nd edition, Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
  - (FML) Foundations of Machine Learning, 2nd edition. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar.

COURSE SCHEDULE

Please see Canvas for the course schedule.

COMMUNICATION

Instructor Communication: All communication from your instructor will take place in Canvas. You are expected to check Canvas every day for important course-related information. However, by following the instructions provided in the course, you can also ensure that you do not miss important instructions, announcements, etc. by adjusting your account settings to receive important information directly to your email account or cell phone. For more details, log into Canvas, enter the course, and see the section entitled "Before You Begin: Instructions for Getting Started."

Communication with TA: Students are assigned to one TA, who is responsible to grade your homework, and answer specific questions and requests. Please find out your assigned TA at the beginning of the semester on Canvas. For all your *administrative requests*, such as regrading homework, please **email your assigned TA, copy to the head TA, and copy to professor.** Your assigned TA will handle all regarding requests.

Content Questions and Help: Because questions can often be addressed for the good of the group, please do not email your questions directly to the instructor. Instead, course and content questions will be addressed on Piazza. Feel free to set your post to private to ask questions about your grade or other issues unique to you. For more information, log into Canvas, enter the course, and see the section entitled "Before You Begin: Instructions for Getting Started." Every day, there will be one TA-on-duty, to answer all question on Piazza on that day. Please allow some time for TAs to respond, especially over the weekend. Please to be courteous when posting on Piazza and treat fellow students, TA, and professor with respect.

Office Hours. Live office hours will be conducted every week via Blue Jeans. These sessions will be both an opportunity for the instructor to discuss course logistics and content but also an opportunity for you to ask questions. While it is strongly suggested that you participate in these meetings, all sessions will be recorded and archived if you are unable to attend or wish to reference them later. For the meeting schedule, links, and archives, please see the section in your Canvas course entitled "Weekly Videoconferences." You may attend any of the office hour that fits into your schedule.

STUDENT EFFORT

Students are expected to devote 8-10 hours per week to complete the course requirements. This guideline encompasses all class activities, including reading the textbook and supplementary resources, watching lesson videos, participating in office hours and forum discussions, completing homework assignments, and studying for exams. Of course, students can spend as much time as necessary, but it is important to be careful not to fall behind.

Assignment Distribution and Grading Scale

Assignment	Release Date	Due Date	Weight
HW1	May 11, 2020	May 24, 2020	(total of 5 HWs) 80%
HW2	May 24, 2020	June 7, 2020	
HW3	June 7, 2020	June 21, 2020	
HW4	June 21, 2020	July 5, 2020	
HW5	July 5, 2020	July 19, 2020	
Project Report	NA	July 29, 2020	20%

GRADING

Grades will be assigned on the following basis:

Homework:	80%
Project:	20%

Important: Make sure the scores in Canvas are consistent with what you got. We will not make any change in grading for works older than 2 weeks. We cannot accept any regrading requests after August 1.

The following grading scale, with scores rounded to the nearest whole number, will be used in the course:

- 90-100%: A
- 80-89%: B
- 70-79%: C
- 60-69%: D
- below 60%: F

LESSONS

Video lessons for this course will be housed on edX. For more details on creating and linking your edX account, log into the Canvas, enter the course, and see the section entitled "Before You Begin: Instructions for Getting Started."

HOMEWORK

Homework should be submitted in Canvas **by 11:59 pm EST on the date it is due**. No submission will be accepted through email. We strongly encourage the use of LaTeX for your submission. Assignments will include both exercises and computer problems; the computer problems will ask you to carry out statistical analysis using computer statistical software. **Keep in mind that you should NOT hand in raw computer output. Conclusions and interpretation of results are more important than good printouts. Compute output with proper explanation will not receive full grades. Please use MATLAB and Python for homework programming.**

*** (Important) Homework code submission requirement**

Please make sure you read through the below guide for homework submission and strictly follow the instruction. For each assignment, two files need to be submitted through Canvas before the deadline:

(1) A pdf with a name: 'Lastname_Firstname_HWx_report'. In this report file, you need to summarize **ALL** your write-up answers to **ALL** questions, including analytical answers, program

output/images/summaries, etc. (Please do not include your program code in the pdf). If Jupyter Notebook is used, you need to print it into the required pdf. For this report, we strongly recommend type-up your answers. You can use either Jupyter Notebook, Latex, or any other mark-down tool. If you choose to hand-writing the report, please make sure your report is clear enough to avoid unnecessary misunderstanding.
(2) A zip file with name "lastname_firstname_HWx.zip": In this zip file, you need to have a folder with the same name: 'lastname_firstname_HWx.', which should include all your program files.

Please make sure you strictly follow the above file format requirements to avoid unnecessary loss of points. In addition, for the programming assignment, you can use either Matlab or Python (2.7 or 3.X). When you submit code, please also include data in the folder. **TA will not be allowed to modify your code.** So please make sure your code doesn't have directory dependence on your computer. If your code has a path issue, it will be considered as "not executable."

You can work together with other students on homework, as long as you write-up and turn in your own solutions. You are also allowed (and encouraged) to ask me questions, although you should try to think about the problems before asking. Request for re-grading the Homework/Exams/Quizzes should be made within a week of returning Homework/Exams/Quizzes. Any kind of academic misconduct is subject to F grade as well as reporting to the Dean of students.

We have the following accommodation policies to help with emergent situations:
(1) You can have TWO CHANCES to ask for a one-week homework extension.
(2) If you have already used the one-chance one-week extension, and if you submit the homework late: one day late the grade will be discount to 75% of your total, two days late the grade will discount to 50% of your total, three days late the grade will discount to 25% of your total. Past three days, your homework will not be accepted.

The regrading requests for both the homework have to be submitted within 2 weeks from the grades are released.

FORMAL EXCUSES
Officially, there are 2 channels a student can get formal excuses. A student can request from Dean's office for illness or other personal issues by filling out a form at https://gatech-advocate.symplicity.com/care_report/index.php/pid122802?. The other way is to request from registrar for institute approved such as attending conference, religious observances (within first 2 weeks) https://registrar.gatech.edu/info/institute-approved-absence-form-for-students.

PROJECT
Project is done by group; each group consists of **1-3** students. Please plan for project early; you need to form your own team (e.g., you can ask on Piazza). You will also need to come up with the topic of your project. By the end of the semester, **July 29, 2020**, each group only needs to submit one project report. Please see Project Guideline for other details.

PLAGIARISM
Plagiarism is considered a serious offense. You are not allowed to copy and paste or submit materials created or published by others, as if you created the materials. All materials submitted and posted must be your own original work.

STUDENT HONOR CODE

You are responsible for completing your own work. All OMS Analytics degree students are expected and required to abide by the *letter* and the *spirit* of the Georgia Tech Honor Code. The teaching assistants and I will also abide by these honor codes. I am very serious about this expectation because ethical behavior is extremely important in all facets of life. To review the Georgia Tech Honor Code, please visit http://osi.gatech.edu/content/honor-code . Any OMS Analytics degree student suspected of behavior in violation of the Georgia Tech Honor Code will be referred to Georgia Tech's Office of Student Integrity.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404) 894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Course Schedule

All assignments and exams are due by 11:59 pm EST on the date listed below.

Week/Dates	Module/Topic	Weekly Overview	Deliverables
Week 1	Introduction and Overview, Clustering and k-means, Spectral Clustering	Overview of the topics and scope of the class. We will introduce a building block of a fundamental problem in unsupervised learning (clustering): K-means. We will discuss another type of clustering algorithm: spectral clustering, which is different from k-means since it is based on geometry (connectivity) of data	Homework 1, release May 11, Due May 24
Week 2	Dimensionality Reduction and PCA, Nonlinear Dimensionality Reduction	We will present linear dimensionality reduction technique called PCA, and non-linear dimensionality reduction techniques	
Week 3	Density Estimation	Discuss basic density estimation method, which captures the distributional information of the data	Homework 2, release May 24, Due June 7
Week 4	Gaussian Mixture Model and EM Algorithm	We will present a popular type of model for densities called Gaussian mixture models and discuss how to fit such models	
Week 5	Basic of Optimization Theory	We will introduction the essentials of optimization theory which is a foundation of developing machine learning algorithms	Homework 3, release June 7, Due June 21
Week 6	Classification Naïve Bayes and Logistic Regression	Introduce classification problem and two basic methods for classification	
Week 7	Support Vector Machine (SVM)	Introduce SVM classifier	Homework 4, release June 21, Due July 5
Week 8	Neural Networks	Understand basic neural networks	
Week 9	Boosting Algorithms and AdaBoost, Random Forest	Introduce basic boosting algorithms and AdaBoost. Introduce tree-based methods for regression and classification, and random forest	Homework 5, release July 5, Due July 19
Week 10	Bias-Variance Tradeoff and Cross-Validation	Introduce principle of bias-variance tradeoff and how it is used to cross-validation for model selection and parameter tuning	
Week 11	Final Week	Summary of class	**Project Report Due –** July 29, 2020.