CSCI 5090/7090 - Machine Learning

Spring 2018

Instructor: Dr. Mehdi Allahyari <mallahyari@georgiasouthern.edu>
Office: CEIT 2321
Course Webpage: https://sci2lab.github.io/mehdi/teaching/x90
Piazza: http://piazza.com/georgiasouthern/spring2018/cscix090

Lecture Times

	Monday	Wednesday
CRN 18045/11783	09:05 AM – 10:20 AM CEIT 2212	09:05 AM – 10:20 AM CEIT 2212

Office Hours

Monday	Wednesday	
11:00 AM – 12:00 PM	11:00 AM – 12:00 PM	
Tuesday	Thursday	or by appointment
03:30 PM – 04:30 PM	03:30 PM – 04:30 PM	
CEIT 2321	CEIT 2321	

Course Description: Machine Learning methods aim to enable computers to automatically learn and improve their performance through experience (e.g., programs that learn to recognize human faces, recommend music and movies, and detect Spam emails). This introductory course is designed to give upperlevel undergraduate and graduate students a broad overview of many concepts and algorithms in ML. The goal is to provide students with a deep understanding of the subject matter and skills to apply these concepts to real world problems. This course covers the theory and practical algorithms for machine learning from a variety of perspectives. We cover topics such as statistical and probabilistic methods, generative and discriminative models, linear and logistic regression, Bayesian networks, decision tree learning, unsupervised learning and clustering and dimensionality reduction. In addition, the course covers fundamental concepts such as training, validation, overfitting, and error rates.

Prerequisites: Students entering the class are expected to have a pre-existing working knowledge of probability, linear algebra, statistics and calculus. For example, it is expected that you know about standard probability distributions and also how to calculate derivatives. In addition, recitation sessions will be held to review some basic concepts. For the programming assignments, you should have good background in programming, and it would be helpful if you know Python.

Textbook: There is no required textbook for this course. There are several recommended books:

- *Machine Learning*, Tom Mitchell
- *Pattern Recognition and Machine Learning*, Christopher Bishop.
- *Machine Learning: A probabilistic perspective*, Kevin Murphy
- *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Trevor Hastie, Robert Tibshirani, Jerome Friedman

Piazza: All students must enroll in the CSCI 5090/7090 course on www.piazza.com. You will receive an enrollment e-mail to your GSU e-mail account before class begins. Follow the directions in the email to set up your account. If you have any questions on how to enroll, please see your instructor. This website will be used for course discussions, announcements, and Q & A. Piazza announcements are considered official course announcements and will be sent to the email address you use for Piazza. If you have a question about anything pertaining to course material, please post the question to piazza before asking your professor during office hours. Typically, many students have the same questions and others will benefit from seeing the Q & A on piazza. The top student answerers on Piazza may receive extra credit toward their final letter grade in the class. *You are NOT allowed to post any code or solutions to labs / assignments on piazza.* This is used only for general Q & A.

Asking Questions: If you have a question about a lab, project, or general course issue, here are the ways to get an answer, rated from best to worst:

1. Search for the answer yourself. Far too often students ask a question whose answer is available on this very page or on the top of assignment handouts.
2. Look for your question on Piazza, then ask a new one if necessary.
3. Ask your TA during lab or office hours.
4. Ask your professor in office hours.
5. Ask your professor in lecture.
6. Send your TA e-mail.
7. Send your professor e-mail.

Course Requirements and Grading: The format of the class will be lecture, with some discussion. I strongly encourage interaction and questions. There are assigned readings for each lecture that are intended to prepare you to participate in the class discussion for that day.

The grading in the class will be divided up as follows:

- Homeworks (30%)
- Midterm Exam (20%), will be given *roughly* in the 8th week of the class
- Final Exam (25%), May 2, 2018 @ 07:30 AM - 09:30 AM
- Final Project (20%)
- Attendance and Participation (5%)

Final Letter Grades: Final letter grades will be determined according to the following scale:

$A \geq 85$	$85 > B \geq 80$	$80 > C \geq 70$	$70 > D \geq 60$	$F < 60$

The grade that you earn in this course will be the grade assigned. There are no bonus assignments and certainly no free points given out at the end (regardless of your reasoning). Bonus points are reserved for the students that answer the most questions on Piazza. Students must be registered for this course in order to receive any grades.

Homework Assignments: The best way to learn about a machine learning method is to program it yourself and experiment with it. Therefore, the assignments will usually involve implementing machine

learning algorithms, and experimentation to test your algorithms on some data. You will be asked to summarize your work, and analyze the results, in brief (3-4 page) write ups. The implementations may be done in Matlab, R or Python, though Python is reocmmended.

Collaboration Policy: Homeworks must be done *individually*, except where otherwise noted in the assignments. 'Individually' means each student must hand in their own answers, and each student must write their own code in the programming part of the assignment. However, students can collaborate in figuring out answers and helping each other solve the problems at **high level**. This means it should not involve any sharing of pseudocode or code or simulation results. Violation of this policy will result in the grade of F, in accordance with university regulations.

Late Penalties: Homework submissions are allowed up to 48 hours after the deadline. *Students turning in a homework after the due date will receive a point deduction as follows:.*

1. If submitted 0–24 hours after the deadline, then 10 points will be deducted from the homework score.
2. If submitted 24–48 hours after the deadline, then 20 points will be deducted from the homework score.
3. If submitted more than 48 hours after the deadline, a score of 0 will be given for the homework.

Email: Email should be used only for personal matters that need to be discussed with the instructor and regrade requests. All questions about course material should be posted to Piazza. Email communication should NOT be treated as an alternative to meeting with the instructor (or TA) during office hours. Email will not be used to provide private tutorials or to explain material that was covered in missed lectures. **Professors and TAs will not be able to debug programs through email**. This must be done during office hours. Students must use their GSU email accounts and put a `[csx90]` tag in the subject of their emails when corresponding with the instructor or TA on course-related matters

In-Class and Online Behavior: Students are expected to be courteous and respectful in all interaction with other members of the class (whether this interaction occurs online or in class). Disruptive or disrespectful behavior might result in the student being asked to leave the classroom. In extreme cases, or if the behavior persists, a formal report might be filed by the instructor or the student withdrawn from the class.

Academic Honesty:

As a Georgia Southern University student, you have agreed to abide by the University's academic honesty policy and the Student Honor Code:

"On my honor, I pledge to be academically honest in all my coursework and will not tolerate the academic dishonesty of others. I also pledge to engage in ethical behavior on-campus and off- campus, to live an honorable lifestyle, and to create a campus environment that is characterized by individual responsibility, civility, and integrity."

Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. Questions related to course assignments and the academic honesty policy should be directed to the instructor. Academic honesty means that any work you submit is your own work.

Common forms of academic dishonesty against which students should guard are:

1. Copying from another student's test paper or laboratory report, or allowing another student to copy from you;
2. Fabricating information, research, and/or results;
3. Collaborating with others on assignments without the faculty's consent;
4. Helping another student to write a laboratory report or computer software code that the student will present as her or his own work, or accepting such help and presenting the work as your own;
5. Cooperating with and/or helping another student to cheat;
6. Turning in material from a public source such as a book or the Internet as your own work.

Steps to help prevent academic dishonesty are:

1. Familiarize yourself with the regulations.
2. If you have any doubt about what constitutes academic dishonesty, ask your instructor or a staff member at the Office of the Vice President for Instruction.
3. Refuse to assist students who want to cheat.
4. Do not allow anyone to copy any of your work, and report anyone who tries to copy from you to the instructor or TA as soon as possible.
5. Do not post your code to a public website. This is a violation of the academic honesty policy as unauthorized assistance

For further information please refer to the Georgia Southern Student Conduct Code, available at the URL below.

http://students.georgiasouthern.edu/conduct/student-code-of-conduct/

Class Accommodation: Students with a disability or health-related issue who need a class accommodation should make an appointment to speak with the instructor or contact Student Disability Resource Center[1] as soon as possible. Students who require such an accommodation for an exam must contact the instructor at least two weeks before the exam is scheduled.

Class Schedule (Tentative): Table below shows the topics of the lectures that will be covered in this course. Please note that all of these are subject to change.

Table 1: My caption

Week	Topic	Note
Jan 08	Course overview	
Jan 15	Decision Trees	
Jan 22	Review of probability	
Jan 29	Probability and Estimation	
Feb 05	Naive Bayes	
Feb 12	Gaussian Naive Bayes	
Feb 19	Linear Regression	
Feb 26	Logistic Regression	
Mar 05	Review and Midterm	
Mar 12	Spring Break	No Classes
Mar 19	Ensemble Methods	
Mar 26	Graphical Models	
Apr 02	Hidden Markov Models	
Apr 09	Clustering	
Apr 16	Dimensinality Reduction	
Apr 23	Neural Networks	

Syllabus Policy: Students are responsible for learning and following all policies stated in this syllabus. This course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

[1] http://students.georgiasouthern.edu/sdrc/