

MSA 8010/CIS 8005: Data Programming

Syllabus for Fall 2016

Contents

1	Description	2
1.1	Sections	2
1.2	Instructor	2
1.3	Contact the instructor	2
1.4	Course Web-site	2
2	Overview	2
2.1	Intended audience	3
2.2	Learning objectives	3
3	Schedule	3
4	Readings by Session	3
4.1	References	5
4.2	E-Books and other resources	5
4.3	Software	5
5	Evaluation	6
6	Homework	6
7	Group Project	6
8	Examinations	7
9	Literature	7
10	Workload Expectations	7
11	Student Behavior	9
11.1	Discrimination and harassment	9
11.2	Official department class policies	9

1 Description

This course introduces common programming practices for data analysis. The objective is to prepare students with programming tools to tackle large data analysis projects.

1.1 Sections

Room: RCB Buckhead Center 306

Days: Thursdays

Time: 7:15 PM – 9:45 PM

1.2 Instructor

Dr. Péter Molnár

Email: pmolnar@gsu.edu

Office: (404) 413-7713

Office hours: TBA & by Appointment.

1.3 Contact the instructor

During the term, it is highly recommended that you contact the instructor, in-person or via email. I am available to help you focus your projects, gain access to resources, and answer your questions. Please try to see me, phone me, or e-mail me at least once during the term to discuss your project. Your class members are also a good source of help.

1.4 Course Web-site

Class information will be posted on the D2L site. There will be links to other web-sites with course related material.

2 Overview

The course introduces

2.1 Intended audience

The course is aimed to data science and analytics students with familiarity with the Python programming language.

2.2 Learning objectives

Upon successful completion of this course, you will accomplish the following objectives and outcomes. In particular, students who complete this course will gain “Ready for work” skills, including:

1. Specifying and reasoning about machine learning problems
2. Applying machine learning tools & techniques

Specific objectives include the following:

1. Understand the machine learning context
2. Understand common machine learning models
3. Understand and apply common mining tools
4. Create descriptive and predictive models from data
5. Know current concepts in machine learning, machine learning, and prediction
6. Demonstrate critical thinking, integrative reasoning, & communication skills

3 Schedule

The course schedule is shown in Table 1. However, the topics and readings may change according to the interests and abilities of the class. See the Academic Calendar. Materials may be updated 24 hours prior to class; please check before attending class.

4 Readings by Session

Readings provide content for class discussions. Thus, readings must be read prior the class. Don’t get more than 1 week ahead of the class in the readings. Sometimes (mostly rarely) readings may be changed 1 week prior to their presentation in class. Table 1 shows the readings by session.

	Date	Topic	Reading	In class	Due
1	Aug 25	Tools, UNIX command line, file system, Python repl, Jupyter notebook etc.			
2	Sep 1	Python in a Nutshell, functionH-ons, classes, data types	(online)		
3	Sep 8	pandas, Plotting	(online)	Lect, H-on, QZ1	HW1
4	Sep 15	numpy, BLAS	(online)	Lect, H-on, QZ2	HW2
5	Sep 22	DataFrames Pandas: group, join	(online)	Lect, H-on, QZ3	HW3
6	Sep 29	Spark, Spark SQL	(online)	Lect, H-on, QZ4	HW4
7	Oct 6	Text processing, regular expressions, NLTK	(online)	Lect, H-on, QZ5	HW5
8	Oct 13	DateTime, Time series	(online)	Lect, H-on, QZ6	HW6
9	Oct 20	Machine Learning Intro, Evaluation, Scikit Learn, Spark ML	ML chapter 2, and 8	Lect, H-on, QZ7	HW7
10	Oct 27	ML Information Based	ML chapter 4	Lect, H-on, QZ8	HW8
11	Nov 3	ML Probability Based	ML chapter 5	Lect, H-on, QZ9	HW9
12	Nov 10	ML Similarity Based	ML chapter 6	Lect, H-on, QZ10	HW10
13	Nov 17	ML Error Based	ML chapter 7	Lect, H-on, QZ11	HW11
	Nov 24	(Thanks Giving Break)			
14	Dec 1	Project		H-on	
15	Dec 8	Presentations		Presentations	

Table 1: Additional reading material will be posted on the web-site. Unless indicated otherwise chapter number refer to the primary text book.

4.1 References

Students must have access to the primary textbook:

Primary Textbook:

- Kelleher, Mac Namee, D'Arcy *Machine Learning for Predictive Data Analytics* MIT Press, 2015.

Readings

1. Murphy, K.P. *Machine Learning: a probabilistic perspective* MIT press, 2012.
2. Mitchell, T.M. *Machine Learning* McGraw-Hill, 1997.

Some books can be accessed from E-book from Books24x7. Most articles have a URL, which can be used to . download the article. (This assumes that you are on the university network directly or VPN. You may be prompted for your campus ID and password.) Some articles may be only available from our web site. To find other articles, use the method described in Section 9.)

4.2 E-Books and other resources

Consider the E-books as good resource; they are free to our students. See this note: <http://www2.cis.gsu.edu/cis/news/newandnoteworthy2.asp>

- Books 24x7. Access from the GSU online library: <http://homer.gsu.edu/search/databases/proxy/GLL25038>; select the link Books24x7
- IT eBooks <http://it-ebooks.info> Select “Search eBook by: Title”. The site hosts complete textbooks for download in PDF.

4.3 Software

All programming activities will be performed on the Cluster using open source software and libraries. In addition the tools are available for Windows, OS X, and Linux. Your laptop should have Python 3 and Jupyter (aka IPython Notebook) installed. Using the Anaconda installation <https://www.continuum.io/downloads> is a good start to have most of the packages available.

Assignment	Percentage
Quizzes (drop 1)	14%
Homework (drop 1)	57%
Project	29%
Total	100%

Table 2: Deliverables and their weights.

Grade	Percentage	Grade	Percentage
A+	≥ 97	C+	≥ 73
A	≥ 90	C	≥ 70
A-	≥ 87	C-	≥ 67
B+	≥ 83	D	≥ 60
B	≥ 80	F	< 60
B-	≥ 77		

Table 3: Breakout depicts how grades will be assigned under this system.

5 Evaluation

Students are evaluated by the deliverables summarized in Table 2.

6 Homework

Homework assignments are the continuation of a hands-on activities in class. Detailed information about the activity and expectation for successful completion are provided with the instructions. See the web site for the most recent and detailed information on these assignments. **Homeworks are individual assignments!**

7 Group Project

The project should showcase some of the methods and tools are introduced in this course. Teams can comprise up to **five** students, and should form within the first few weeks of the term. Teams are free to choose a data set for their project. Teams should start early on

developing their project proposal. The use of proprietary or classified data sets is not allowed. Project deliverables include a detailed report, functioning code, and a poster. Details about requirements and evaluation criteria will be posted on the web-site.

The group is presenting a digital poster, i.e. a single PowerPoint on the Christie Visualization Wall. The group can elect one or more speakers... please ensure smooth transition between speakers. Do not include any sound or video files. Do not use any animation or transition effects. You may include animated GIFs to visualize effects that change over time (see <https://www.gapminder.org/world/>).

8 Examinations

Quizzes are usually given out in the middle of the class and comprise only a few questions. However, some questions may need some thinking and calculations.

9 Literature

Search for peer reviewed articles using keywords:

1. Scan the web (in particular using scholar search engines)

- (a) <http://scholar.google.com/>
- (b) <http://academic.live.com/>
- (c) <http://citeseer.ist.psu.edu/>

2. Scan using library databases (@GSU)

- (a) <http://www.galileo.usg.edu>
- (b) In particular, the following databases
 - i. ABI/INFORM Complete
 - ii. ACM Digital Library
 - iii. IEEE Xplore

10 Workload Expectations

Students should plan for 2 - 3 hours of work outside of class each week for each course credit hour. Thus, a 3-credit course averages between 6 and 9 hours of student work outside of the

classroom, each week. See GSU sites for Academic Success:

- <http://www2.gsu.edu/~wwwcam/incept/successtips.html>
- <http://www2.gsu.edu/~wwwctr/sac/StudySkills.htm>

Self-Managed Teams: Teams will be allowed for some activities during the term. Please note that unless the activity is explicitly identified as a “team activity”, I expect everyone to perform their own work (your hands on the keyboard). For team activities, you will be allowed to work with partners (of your choosing).

- Initial teams must be established by the second week of classes. Established teams may continue working together on subsequent team activities. Team membership may change during the term, if problems arise. However, team members must be designated within one week of the due date for the team activity. Exception: you may withdraw from a team at any time and submit an assignment individually.
- Teams will submit one assignment for all team members. In most cases, each member of the team will get the same score. However, an individual’s score may be reduced at the discretion of the instructor.
- Each team assignment must include the following:
 - Tasks completed by each member.
 - Percentage of the total work completed by each member.
- Any individual with a low team contribution will be removed from their team.

Arbitration: There will be a one-week arbitration period after graded activities are returned. Within that one-week period, you are encouraged to discuss any assumptions and/or misinterpretations that you made on the activity that may have influenced your grade.

Attendance: If you are unable to attend a class session, it is your responsibility to acquire the class notes, assignments, announcements, etc. from a classmate. The instructor will not give private lectures for those that miss class.

Submission of Deliverables: Unless specific, prior approval is obtained, no deliverable will be accepted after the specified due date. If you have a legitimate personal emergency (e.g., health problem) that may impair your ability to submit a deliverable on time, you must take the initiative to contact the instructor before the due date/time (or as soon after your emergency as possible) to communicate the situation. Make-up exams will not be given: However, if a student has a planned absence, he or she may take the exam earlier with the permission of the instructor.

11 Student Behavior

Behavior in class should be professional at all times. People must treat each other with dignity and respect in order for scholarship to thrive. Behaviors that are disruptive to learning will not be tolerated and may be referred to the Office of the Dean of Students for disciplinary action.

11.1 Discrimination and harassment

Discrimination and/or harassment will not be tolerated in the classroom. In most cases, discrimination and/or harassment violates Federal and State laws and/or University Policies and Regulations. Intentional discrimination and/or harassment will be referred to the Affirmative Action Office and dealt with in accordance with the appropriate rules and regulations. Unintentional discrimination and/or harassment is just as damaging to the offended party. But, it usually results from people not understanding the impact of their remarks or actions on others, or insensitivity to the feelings of others. We must all strive to work together to create a positive learning environment. This means that each individual should be sensitive to the feelings of others, and tolerant of the remarks and actions of others. If you find the remarks and actions of another individual to be offensive, please bring it to their attention. If you believe those remarks and actions constitute intentional discrimination and/or harassment, please bring it to my attention.

11.2 Official department class policies

1. Prerequisites are strictly enforced. Students failing to complete any of the prerequisites with a grade of “C” or higher will be administratively withdrawn from this course with loss of tuition fees. There are no exceptions, except as granted by the instructor with the approval of the department.
2. Students are expected to attend all classes and group meetings, except when precluded by emergencies, religious holidays, or bona fide extenuating circumstances.
3. Students who, for non-academic reasons beyond their control, are unable to meet the full requirements of the course should notify the instructor, by email, as soon as this is known and prior to the class meeting. Incompletes may be given if a student has ONE AND ONLY ONE outstanding assignment.
4. A “W” grade will be assigned if a student withdraws before mid-semester if (and only if) he/she has maintained a passing grade up to the point of withdrawal. Withdrawals after the mid-semester date will result in a grade of “WF”. See the GSU catalog or registrar’s office for details.

5. Spirited class participation is encouraged and informed discussion in class is expected. This requires completing readings and assignments before class.
6. All exams and individual assignments are to be completed by the student alone with no help from any other person.
7. Collaboration within groups is encouraged for project work. However, collaboration between project groups will be considered cheating.
8. Copying work from the Internet without a proper reference is considered plagiarism and subject to disciplinary action as delineated in the GSU Student Handbook.
9. Any non-authorized collaboration will be considered cheating and the student(s) involved will have an Academic Dishonesty charge completed by the instructor and placed on file in the Dean's office and the CIS Department. All instructors regardless of the type of assignment will apply this Academic Dishonesty policy equally to all students. Abstracted from GSU's Student Handbook Student Code of Conduct "Policy on Academic Honesty and Procedures for Resolving Matters of Academic Honesty"

(a) http://www2.gsu.edu/%7Ewwwdos/codeofconduct_conpol.html

(b) <http://www2.gsu.edu/~wwwcam/>

As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. The University assumes as a basic and minimum standard of conduct in academic matters that students be honest and that they submit for credit only the products of their own efforts. Both the ideals of scholarship and the need for fairness require that all dishonest work be rejected as a basis for academic credit. They also require that students refrain from any and all forms of dishonorable or unethical conduct related to their academic work.

Students are expected to discuss with faculty the expectations regarding course assignments and standards of conduct. Here are some examples and definitions that clarify the standards by which academic honesty and academically honorable conduct are judged at GSU.

Plagiarism. Plagiarism is presenting another person's work as one's own. Plagiarism includes any paraphrasing or summarizing of the works of another person without acknowledgment, including the submitting of another student's work as one's own. Plagiarism frequently involves a failure to acknowledge in the text, notes, or footnotes the quotation of the paragraphs, sentences, or even a few phrases written or spoken by someone else. The submission of research or completed papers or projects by someone else is plagiarism, as is the unacknowledged use of research sources gathered by someone else when that use is specifically forbidden by the faculty member. Failure to indicate the extent and nature of one's reliance on other sources is also a form of plagiarism. Any work, in whole or part, taken from the Internet or other computer based resource without properly referencing the source (for example, the URL) is considered plagiarism. A complete reference is required in

order that all parties may locate and view the original source. Finally, there may be forms of plagiarism that are unique to an individual discipline or course, examples of which should be provided in advance by the faculty member. The student is responsible for understanding the legitimate use of sources, the appropriate ways of acknowledging academic, scholarly or creative indebtedness, and the consequences of violating this responsibility.

Cheating on Examinations. Cheating on examinations involves giving or receiving unauthorized help before, during, or after an examination. Examples of unauthorized help include the use of notes, texts, or "crib sheets" during an examination (unless specifically approved by the faculty member), or sharing information with another student during an examination (unless specifically approved by the faculty member). Other examples include intentionally allowing another student to view one's own examination and collaboration before or after an examination if such collaboration is specifically forbidden by the faculty member. Unauthorized Collaboration. Submission for academic credit of a work product, or a part thereof, represented as its being one's own effort, which has been developed in substantial collaboration with another person or source or with a computer-based resource is a violation of academic honesty. It is also a violation of academic honesty knowingly to provide such assistance. Collaborative work specifically authorized by a faculty member is allowed.

Falsification. It is a violation of academic honesty to misrepresent material or fabricate information in an academic exercise, assignment or proceeding (e.g., false or misleading citation of sources, the falsification of the results of experiments or of computer data, false or misleading information in an academic context in order to gain an unfair advantage).

Multiple Submissions. It is a violation of academic honesty to submit substantial portions of the same work for credit more than once without the explicit consent of the faculty member(s) to whom the material is submitted for additional credit. In cases in which there is a natural development of research or knowledge in a sequence of courses, use of prior work may be desirable, even required; however the student is responsible for indicating in writing, as a part of such use, that the current work submitted for credit is cumulative in nature.