MAT 395/495: Scientific Data Analysis and Computing

Syllabus for Summer Semester 2020

Catalog Description

Introduces basic methods and code development tools for scientific computing in the Python language, including coding, analysis, data fitting, visualization, importing/exporting information, multidimensional analysis, and visualization using Python and its packages MatPlotlib, NumPy, SciPy. Basic statistical methods, distributions, data fitting and error analysis will be discussed with a focus on relevancy for material scientists and engineers. In addition to numerical approaches, symbolic problem solving will be discussed. Integration of external data sets, databases queries, and collective tools into computational methods are introduced.

Prerequisites: None

Instructors

Prof. Nick Strandwitz: strand@lehigh.edu
Prof. Siddha Pimputkar: siddha@lehigh.edu
Prof. Joshua Agar: joshua.agar@lehigh.edu

Office Hours

This class will not have dedicated office hours, rather we will handle “office hour” related questions via *Slack*. Traditional office hours may be made by appointment with any of the three instructors should you wish to discuss a topic which is not suitable for Slack.

Slack Workspace: LehighMSE-Classes
Class Channel: #mat395-495-questions
Invite to Workspace Link:
https://join.slack.com/t/lehighmse-classes/shared_invite/zt-f1xei6fm-SIAjuyBHRHcRLQQ77QhkvA

Class Time & Location

Semester: June 29 -- August 6, 2020;

Days: Monday, Tuesday, Wednesday, Thursday
Time: 9:00—10:35 am ET (live lectures)

All classes will be *taught remotely via Zoom* for both synchronous (live) and asynchronous (recorded) content.

Class Platform

This course makes use of multiple digital platforms as detailed below.

Platform	Purpose
Colab (https://colab.research.google.com)	• *Default Python platform* • Assignments and homework *must be* submitted from this platform • Project *may be* performed on any Python platform of your choosing
Google Drive	• Every student has a folder which the instructors will share with them • Submission of Assignments/Homework in student folder
Course site (https://coursesite.lehigh.edu)	• Grades • Syllabus • Links to Zoom Meetings • Lecture Recordings
GitHub (Read-only for students) https://github.com/jagar2/Summer_2020_MAT-395-495_Scientific-Data-Analysis-and-Computing	• Lectures notes • Datasets • Posting of Assignment/Homework
Slack (https://lehighmse-classes.slack.com/)	• General announcements *from instructors* for class (#mat395-495-announcements) • Posting of any questions for class, assignment, homework and answering by both students and faculty (#mat395-495-questions)

Resources for this Class

There is no formal textbook for this course, though one will be used for the statistical analysis section. All software used in this course will be open source. There are many excellent resources including YouTube videos, online tutorials, and books that are available for free. All of the course material will be designed to run on Google Collaboratory a free open-source python instance in the cloud. To access colab you need a computer with internet access. All assignments will be required to be completed in Colab. The final project can be completed on any python interface, however, it must be sufficiently documented to be reproducible by your peers and/or your instructor.

Throughout the course we will make materials available to you on the code hosting platform GitHub. Github allows you to download, modify, and merge changes to code using version control. There are many good tutorials available here is one good example https://www.youtube.com/watch?v=77W2JSL7-r8.

We recommend that you create a fork of the repository. This provides you with your own copy of the repository that you can modify as you like. Git will, however, preserve the original repository in case you ever want to revert the entire or part of the repository. As we update the content on this repository you will be able to merge these changes with your fork.

Throughout the course we will be using a variety of digital resources. All of these resources have extensive free educational materials. Here we provide links to some of the quickstart guides.

Google Collaboratory: Cloud Python Instance
https://www.youtube.com/watch?v=inN8seMm7UI
https://colab.research.google.com/notebooks/intro.ipynb

GitHub: Code-hosting platform with version control
https://guides.github.com/activities/hello-world/
https://www.youtube.com/watch?v=noZnOSpcjYY
https://www.youtube.com/watch?v=FyfwLX4HAxM

Textbook:
No textbook needs to be purchased for this class as relevant information will be provided electronically to the students. There are several additional texts, many good tutorials, and problems solved that can be found on the internet. We will list a few particularly good resources here.

Statistical analysis section discussion will be based on:

- "*Data Reduction and Error Analysis for Physical Sciences*" by P.R. Bevington, D. K. Robinson, 3rd Edition, McGrawHill, 2003, ISBN-13: 978-0-07-247227-1
- "*An Introduction to Error Analysis--The Study of Uncertainties in Physical Measurements*", by John R. Taylor, University Science Books, 1997, ISBN: 0-935702-75-X

Open Course Materials:
Other open course materials for scientific computing in python.
http://sbu-python-class.github.io/python-science/
http://web.stanford.edu/class/cme193/syllabus.html

Topics/Schedule (*Tentative*)

Lecture	Instructor	Topic	Deliverable	Due (after class)
1	NS	Intro to python, github, mechanisms for course, syllabus	Assignment	21 hr
2	NS	More intro to python: variables and syntax	Assignment	21 hr
3	NS	More intro to python	Assignment	21 hr
4	NS	Numpy	Assignment	21 hr
5	NS	Scipy and Sympy (intro)	Assignment	21 hr
6	NS	Pandas	Assignment	21 hr
7	JA	Static Data Visualization: Matplotlib	Assignment HW Assignment	21 Hr 1 Week
8	All	Working Class		
9	JA	Static Data Visualization/Interactive Graphing		
10	JA	Image Analysis: Loading images and image representations, Local and global filters and transformations	Assignment Due	
11	JA	Image Analysis: Loading images and image representations, Local and global filters and transformations	Assignment	21 hr

12	JA	Image Analysis: Loading images and image representations, Local and global filters and transformations	Homework 2	1 Week
13	JA	Image Filter and Segmentation	Assignment	21 Hr
14	SP	Uncertainties in Measurements Error analysis: Types, Propagation	Assignment	21 Hr
15	SP	Understanding distributions: Binomial, Poisson, Gaussian/Normal, Lorentzian, Voigt	Assignment HW	21 Hr 1 week
16	SP			
17	SP	Estimating Mean and Errors: Least squares, Probability tests (T-test, Chi-square)	Assignment	21 hr
18	SP	Least Square Fitting: Straight	Assignment	21 hr
19	SP	Testing the fit: Chi-squared test	Assignment HW	21 hr 1 week
20	JA	Sympy	Assignment	21 hr
21	JA	Linear Machine Learning: classification, regression models, regularizers, and overfitting	HW	1 Week
22	*TBD*	*Class Determined Topic*		
23	--	Project Presentations		
24	--	Project Presentations		

Learning Objectives

At the end of the class, students should be able to achieve the following:

- **Ability to use Python to**:
  - Write simple program
  - Understand use of appropriate data structures
  - Calculate statistical data (errors, std dev, …)
  - Multidimensional analysis
  - Visualize data (line, xy, xyz, etc)
  - Import/Export data / images
- **Statistical Analysis**:
  - Understand and fit distributions to data sets
  - Fit trend lines to data allowing for model development

    - Understand and determine uncertainties and errors in data
- **Symbolic Solving**
    - Solve symbolic equations
    - Perform operations on symbolic equations
- **Advanced topics**:
    - GUI development
    - Interface with public data servers (materialsproject.org, etc.)
    - Numeric analysis algorithm

Structure of Class and Expectations

Lectures: Live lectures will be held and recorded during the regular lecture period. Recordings will be made available after the lecture on course site. Class-periods will be a mix of lecturing with in-class active programming and working on smaller assignments. Lecture periods are used to provide new information and aid in the process of understanding and applying this knowledge to help solve problems. Lectures will be delivered via lecturing by the main instructor on Zoom with aided by visual/digital elements and slides provided on the main screen or course site. In-class activities will be performed as helpful to understanding the subject matter. Active participation during class period is encouraged. Students are expected to prepare for every lecture by reading the appropriate course material and textbook before the class period.

Assignments: Assignments will be assigned after most lectures and will be due before the next lecture. These assignments will be short (~ 30 min) and be more methodological in nature. They are designed to enforce concepts delivered in class and be straight forward.

Homework: Homework will be assigned once a week and be larger in scope. They are more open-ended and involve connecting multiple new concepts learned throughout lecture periods. This will train you in solving open-ended problems with less immediate guidance provided.

Project will be assigned to build on knowledge gained in the individual assignments and lectures, though will be larger in scope than the more targeted and structured assignments. The project should be written using Python.

Assignment/Homework Expectations

Assignments and homework may be solved in a group setting, though submission and work need to be your own as an individual. If you worked in a group, indicate so on your assignment as to avoid the potential problem of plagiarism. Each assignment should be submitted as a professionally structured and written digital file. The assignments and homework are to be submitted via placement in your assigned google drive folder. Data files or other information needed to complete assignments will be made available on GitHub.

Project Expectations

The project will be a personalized assignment focusing on a project of interest to the student's research field. The student's project topic will be developed via consultation with the instructors of this course. The project should be selected by the 5th lecture and prospective projects should be discussed with one or more of the professors by the 2nd lecture. Early project selection is important to give ample time to accomplish project goals. A topic can be provided if the student does not have any research project from which data can be utilized for this course. You will be expected to present your project to the class.

Grading Scheme

The following deliverables contribute to the final grade and will be weighed as follows:

Assignments	30%
Homework	40%
Project	30%
Total	**100%**

Conversion to a letter grade will occur by setting the average of the class to a letter grade representative of the overall performance of the class.

Late submissions: Late submissions on any assignment will be accepted at the sole discretion of the instructor. *By default, a late submission will not be accepted and receive no credit (i.e. a zero grade) unless prior arrangements were made with the instructor, in writing, prior to the submission day and time*. Late submissions will be accepted if accompanied by a letter from the Dean of Students providing justifiable cause for the tardiness (illness or another emergency).

Principles of Our Equitable Community

Lehigh University endorses The Principles of Our Equitable Community (http://www.lehigh.edu/~inprv/initiatives/PrinciplesEquity_Sheet_v2_032212.pdf). We expect each member of this class to acknowledge and practice these Principles. Respect for each other and for differing viewpoints is a vital component of the learning environment inside and outside the classroom.

Accommodations for Students

If you have a disability for which you are or may be requesting accommodations, please contact both your instructor and the Office of Academic Support Services as early as possible in the semester. You must have documentation from the Academic Support Services office before accommodations can be granted.

For more information or to request services, please contact **Disability Support Services**:

Office: Williams Hall, Room 301
Phone: 610-758-4152
Email: indss@lehigh.edu
Website: https://studentaffairs.lehigh.edu/disabilities

Academic Integrity

Lehigh University Student Senate Statement on Academic Integrity:

> *We, the Lehigh University Student Senate, as the standing representative body of all undergraduates, reaffirm the duty and obligation of students to meet and uphold the highest principles and values of personal, moral and ethical conduct. As partners in our educational community, both students and faculty share the responsibility for promoting and helping to ensure an environment of academic integrity. As such, each student is expected to complete all academic course work in accordance to the standards set forth by the faculty and in compliance with the University's Code of Conduct.*

The work you do in this course must be your own. This means that you must be aware when you are building on someone else's ideas—including the ideas of your classmates, your professor, and the authors you read—and explicitly acknowledge when you are doing so. Feel free to build on, react to, criticize, and analyze the ideas of others but, when you do, make it known whose ideas you are working with.

Resources: https://studentaffairs.lehigh.edu/content/academic-integrity-resources
Plagiarism: http://libraryguides.lehigh.edu/plagiarism
Vignettes: http://www.lehigh.edu/lts/official/Academic_Integrity_Vignettes.pdf

Policy on Harassment and Non-Discrimination

Lehigh University upholds The Principles of Our Equitable Community and is committed to providing an educational, working, co-curricular, social, and living environment for all students,

staff, faculty, trustees, contract workers, and visitors that is free from harassment and discrimination on the basis of age, color, disability, gender identity or expression, genetic information, marital or familial status, national or ethnic origin, race, religion, sex, sexual orientation, or veteran status. Such harassment or discrimination is unacceptable behavior and will not be tolerated. The University strongly encourages (and, depending upon the circumstances, may require) students, faculty, staff or visitors who experience or witness harassment or discrimination, or have information about harassment or discrimination in University programs or activities, to immediately report such conduct.

If you have questions about Lehigh's Policy on Harassment and Non-Discrimination or need to report harassment or discrimination, contact the **Equal Opportunity Compliance Coordinator**:

Office: Alumni Memorial Building
Phone: 610-758-3535
Email: eocc@lehigh.edu
Website: https://eocc.lehigh.edu/