SYSTEMS DEVELOPMENT FOR COMPUTATIONAL SCIENCE

CS107 / AC207

Fabian Wermelinger
Harvard University

Tuesday, August 16th 2022

COMPUTATIONAL SCIENCE AND LEARNING OBJECTIVES

Computational Science and Engineering (CSE) can be seen as a third mode of discovery next to theory and experiment. Computer simulation is what drives CSE and an important component in research and industry today. *This class teaches the necessary requirements for successful development of software systems that are the foundation of computer simulation.*

After successful completion of this course, the student will be able to:

- Use Python, including its advanced features to write scientific programs.
- Understand the Python data model and how it impacts the code you write.
- Write programs with good software engineering practices. These practices include: working on remote machines, version control, continuous integration, documentation and testing.
- Utilize data management techniques to store data, starting from a good understanding of data structures to databases.

FOR WHOM IS THIS CLASS?

- You want to learn how to automate workflows using (BASH) scripts and control versioning of the code you write.
- You want to deepen your knowledge in programming with Python and learn its underlying data model.
- Learn about containerization, continuous integration, testing, documentation and deployment of software projects.
- Learn the basics of data structures and databases.

CS107/AC207 is an *interdisciplinary* class and welcomes students from all fields in Science, Math, Engineering and Technology.

Prerequisites:

- You should have some basic familiarity with programming (functions, variables, constants, etc.) at the level of CS50.
- You should have basic calculus knowledge (chain rule). The lectures will review the necessary fundamentals required to succeed with the class project.
- You should have interest and joy for scientific computing (not a strict requirement but it helps).
- Download Homework 0 for self-assessment (you do not need to be able to solve all problems in order to take this class).

CLASS COMPONENTS

Main class website: https://harvard-iacs.github.io/2022-CS107/

- ***Canvas:*** main hub for grades, class discussion forum and other sensitive content.
- ***Class Git repository:*** all class handouts are provided via Git. You submit your assignment solutions via your private Git repository as well.
- ***Discussion forum:*** we use an EdStem discussion board that has been installed on canvas. The platform is used for augmented discussion of lecture topics and assignments.

Grading:

Homework:	35%
Project:	35%
Quizzes:	15%
Pair-programming:	10%
Communal Contributions: Via the class communication platforms.	5%

- CS107/AC207 does not have standard midterm or final exams.
- The project work involves presentations.
- Quizzes are out-of-class via canvas, open book and `www` accessible.

SCHEDULE

Legend: shell scripting, version control, CI, development tools, Python, Python (advanced), data structures and databases.

Wk	Tuesday	Thursday	Labs	Events
1(35)		***Lecture 1:*** *2022-09-01* • Class introduction/organization • History of Bell Labs, Unix and Linux • Command line introduction	***Sign-up:*** Select one of the offered pair-programming lab session days according to your schedule ***PP1:*** *(2022-09-02)* Setup private class repository, `tmux`	***Note:*** Handouts are typeset in green and deadlines in red. All deadlines are due 11:59 pm. 1. HW1 release (2022-09-01) 2. Doodle for pair-programming day due (2022-09-02)
2(36)	***Lecture 2:*** *2022-09-06* • Manual pages and help for Linux commands • Unix philosophy • Regular expressions and `grep` • File attributes and permissions • Short and unbiased journey into text editors	***Lecture 3:*** *2022-09-08* • Shell customization • I/O redirection • Process management • Environment variables and shell scripting	***PP2:*** *(2022-09-09)* Git workflow and `bash` scripting	
3(37)	***Lecture 4:*** *2022-09-13* • Introduction to version control systems (VCS) • Centralized and distributed VCS • Design and inner workings of Git	***Lecture 5:*** *2022-09-15* • Git blobs, trees and commits • Git status of working tree and index • Git remote repositories • Git branches	***PP3:*** *(2022-09-16)* Git branches and merge conflicts	1. HW2 release (2022-09-13) 2. HW1 due (2022-09-13) 3. PP1 due (2022-09-16)

SCHEDULE

Legend: shell scripting, version control, CI, development tools, Python, Python (advanced), data structures and databases.

Wk	Tuesday	Thursday	Labs	Events
4(38)	***Lecture 6:*** *2022-09-20* • Introduction to Python • Nested environments and closures • Decorators	***Lecture 7:*** *2022-09-22* • Object Oriented Programming (OOP) • Classes in Python • Inheritance and polymorphism • ***Quiz 1***	***PP4:*** *(2022-09-23)* Fully connected neural networks and Python closures	1. Project M1A due (2022-09-22) 2. PP2 due (2022-09-23) 3. C/C++ primer class 09/19 - 09/23
5(39)	***Lecture 8:*** *2022-09-27* • Duck typing • The Python data model • Special methods (dunder methods) • *Aside:* software licenses and open source	***Lecture 9:*** *2022-09-29* • Python class methods, static methods, instance methods • Python class attributes and instance attributes • Python modules • Python packages and PyPI	***PP5:*** *(2022-09-30)* Python classes and dunder methods	1. HW3 release (2022-09-27) 2. HW2 due (2022-09-27) 3. PP3 due (2022-09-30) 4. C/C++ primer class 09/26 - 09/30
6(40)	***Lecture 10:*** *2022-10-04* • Preliminary automatic differentiation (AD) • Derivatives and the Jacobian • Newton's method • Numerical approximation of derivatives (finite-difference method)	***Lecture 11:*** *2022-10-06* • Review of chain rule • Evaluation trace of a function and computational graph • Forward mode AD • Forward mode in higher dimensions	***PP6:*** *(2022-10-07)* Forward mode automatic differentiation	1. Project M1B due (2022-10-04) 2. PP4 due (2022-10-07)

SCHEDULE

Legend: shell scripting, version control, CI, development tools, Python, Python (advanced), data structures and databases.

Wk	Tuesday	Thursday	Labs	Events
7(41)	***Lecture 12:*** *2022-10-11* • Dual numbers and complex numbers • Implementation of forward mode AD: operator overloading	***Lecture 13:*** *2022-10-13* • Reverse mode AD • Some examples AD applications and extensions • ***Quiz 2***	***PP7:*** *(2022-10-14)* Forward mode AD, Jacobian, seed vector	1. HW4 release (2022-10-11) 2. HW3 due (2022-10-11) 3. PP5 due (2022-10-14)
8(42)	***Lecture 14:*** *2022-10-18* • Continuous integration (CI) • Testing Python code (`pytest` and `unittest`) • Code coverage • Documenting code (docstrings, shpinx, readthedocs)	***Lecture 15:*** *2022-10-20* • Python virtual environments • Introduction to docker containers and `Dockerfiles`	***PP8:*** *(2022-10-21)* Python virtual environments and deploying Python packages	1. Project M1 due (2022-10-20) 2. PP6 due (2022-10-21)
9(43)	***Lecture 16:*** *2022-10-25* • Building your own docker containers • Integration of custom containers in CI workflows	***Lecture 17:*** *2022-10-27* • Abstract data types • Linked lists • Iterators	***PP9:*** *(2022-10-28)* Binary trees, binary search trees	1. HW5 release (2022-10-25) 2. HW4 due (2022-10-25) 3. PP7 due (2022-10-28)
10(44)	***Lecture 18:*** *2022-11-01* • Trees and binary trees • Binary search trees (BST) • Priority queues and heaps	***Lecture 19:*** *2022-11-03* • Python generators • Python coroutines • Introduction to Python internals	***PP10:*** *(2022-11-04)* Python generators	1. Project M2A due (2022-11-01) 2. PP8 due (2022-11-04)

SCHEDULE

Legend: shell scripting, version control, CI, development tools, Python, Python (advanced), data structures and databases.

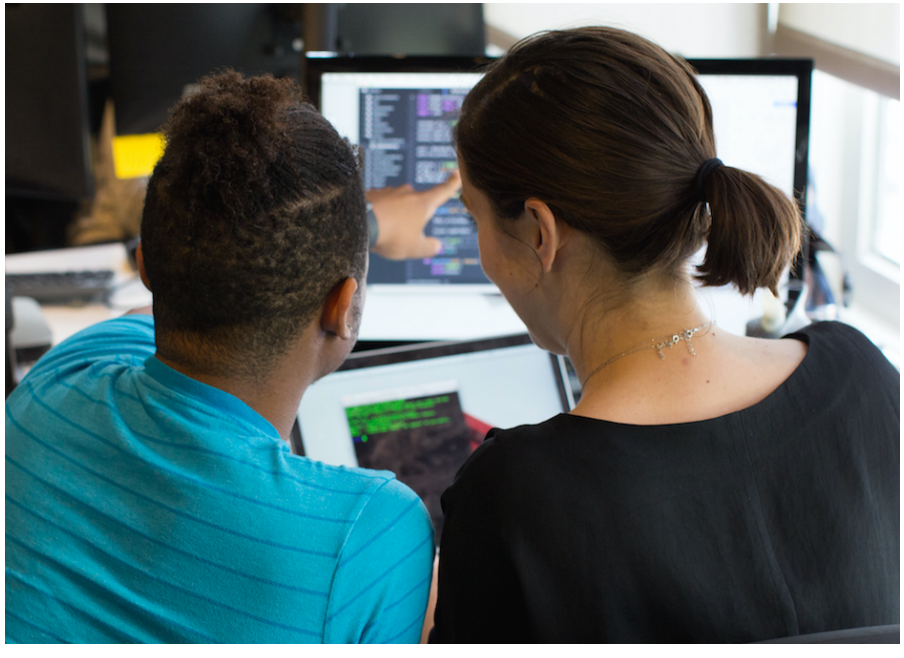
Wk	Tuesday	Thursday	Labs	Events
11(45)	***Lecture 20:*** *2022-11-08* • Code objects and Python bytecode • The Python interpreter and the evaluation loop • Memory allocation in Python • Performance of pure Python lists and NumPy arrays	***Lecture 21:*** *2022-11-10* • Data models and databases • Structured query language (SQL) • SQLite in Python • ***Quiz 3***	***PP11:*** *(2022-11-11)* Databases	1. HW6 release (2022-11-08) 2. HW5 due (2022-11-08) 3. Project M2B due (2022-11-10) 4. PP9 due (2022-11-11)
12(46)	***Lecture 22:*** *2022-11-15* • Databases • In-class exercise with SQL and Python SQLite	***Lecture 23:*** *2022-11-17* • Databases • In-class exercise with SQL and Python SQLite • Table joins • Pandas	***PP12:*** *(2022-11-18)* Finish in-class database exercises	1. Project M2 due (2022-11-17) 2. PP10 due (2022-11-18)
13(47)	***Lecture 24:*** *2022-11-22* • Debugging in Python • Generating profiles for performance analysis • Bytecode instructions and performance	***Thanksgiving break:*** *2022-11-24*		1. HW7 release (2022-11-22) 2. HW6 due (2022-11-22) 3. PP11 due (2022-11-25)

SCHEDULE

Legend: shell scripting, version control, CI, development tools, Python, Python (advanced), data structures and databases.

Wk	Tuesday	Thursday	Labs	Events
14(48)	***Lecture 25:*** *2022-11-29* • Project work • *Quiz 4*	***Lecture 26:*** *2022-12-01* • Project work		1. PP12 due (2022-12-02) 2. HW7 due (2022-12-04)
15(49)	***Reading period:*** *2022-12-06*	***Exam period:*** *2022-12-08*		1. Project final milestone due (2022-12-10)
16(50)	***Exam period:*** *2022-12-13*	***Exam period:*** *2022-12-15*		

LABS / PAIR-PROGRAMMING



https://harvard-iacs.github.io/2022-CS107/pages/syllabus.html#pp

- One section per week is a pair-programming exercise lead by TFs.
- These sections are graded based on attendance and completion.
- You can choose a section day that best fits your schedule at the beginning of the semester.
- Exercises support lecture and homework content and help you further *practice* the material.
- Learn new command line tools that allow you to share interactive sessions. These tools are also great for interactive debugging sessions for the project.

FINAL PROJECT

- A substantial part of the grade in CS107/AC207 consists of a group project.
- The topic of the project is automatic differentiation (AD), a method to compute and evaluate derivatives of arbitrary functions up to machine precision.
- Your task is to develop an AD library in Python that involves the software development techniques discussed in class.
- You work in groups of 4-5 students (you can choose your partners).
- The project runs side-by-side with the class and is divided into 7 milestones in total (see the schedule for deadlines).
- The lectures will review the basic calculus required for AD and teach the *forward* and *reverse* modes of AD in lectures 11 to 13.
- See Homework 0 for some basic calculus exercises.

https://harvard-iacs.github.io/2022-CS107/pages/project.html

C/C++ PRIMER CLASS (OPTIONAL)

- You can join a voluntary C/C++ primer class (recommended if you plan on taking CS205).
- The class is intended to strengthen basic C/C++ programming *(built-in types, arrays, pointers, memory allocation, operators, functions, object oriented design, inheritance and polymorphism, operator overloading, generic programming, STL and writing C++ modules for Python)*.
- Some basic knowledge of compiled programming languages is helpful to fully benefit.
- ***Everybody is welcome!*** (even if you are not taking CS107/AC207 or intending to take CS205). *If you know somebody who you think can benefit from this primer, please invite them*.

  - 10 lectures at 75 minutes each
  - ***September 19 to 30 at 5:30pm - 6:45pm weekdays*** in SEC LL2.224 SU Family Classroom

  See https://harvard-iacs.github.io/2022-CS107/pages/cpp_primer.html for details and registration.

CS107/AC207 CLASS WEBSITE

All information available on class website:
https://harvard-iacs.github.io/2022-CS107/

(You can find these slides posted on the main class website in the Updates section.)

OFFICE HOUR THIS WEEK

For further questions about CS107/AC207 please come see me in the office hour on ***Thursday August 18th 10:00am - 11:00am (EST)*** on zoom.

Happy to answer questions now as well if time allows.