Course Syllabus

18-847F: Special Topics in Computer Systems
Foundations of Cloud and Machine Learning Infrastructure
Fall 2019

Class Hours: Mon and Wed, 4:30 – 6:00 pm
Class Location: Scaife Hall 222

Number of Units: 12
Pre-requisites: Introductory probability, linear algebra

Instructor: Prof. Gauri Joshi (gaurij@andrew.cmu.edu)
Office Location: CIC 4105
Office Hours: By appointment

Teaching Assistants:
Abhishek Sawarkar (asawarka@andrew.cmu.edu)
Jianyu Wang (jianyuw1@andrew.cmu.edu)

Course Description: The objective of this graduate-level seminar course is to introduce students to algorithms in large-scale computing and machine learning. You will read, present and critique a curated set of research papers from both theory and systems. Each class will comprise of presentation and discussion of two research papers. The first half of the course will cover distributed computing frameworks and scheduling and load balancing policies used in them. In the context of distributed storage, we will discuss coding-theoretic techniques used to improve availability and repair failed nodes. The second half of the course will focus on machine learning infrastructure. We will cover distributed SGD, federated learning and hyper-parameter tuning.

Course Website: https://www.andrew.cmu.edu/course/18-847F/. The latest course calendar, along with the list of upcoming speakers will be posted here.

Reference Textbooks (supplementary reading; not necessary for lectures):
1. *Understanding Machine Learning: From Theory to Algorithms,* Shai Ben-David and Shai Shalev-Shwartz, Cambridge University Press, 2014
2. *Convex Optimization*, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, 2004
3. *Deep Learning,* Ian Goodfellow, Yoshua Bengio and Aaron Cornville, MIT Press, 2016, http://www.deeplearning.org
4. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*, Mor Harchol-Balter, Cambridge University Press, Feb 2013.

5. *Bandit Algorithms*, Tor Lattimore and Csaba Szepesvari, http://downloads.tor-lattimore.com/banditbook/book.pdf, Aug 2018

Homework:
Each homework assignment will consist of concept-check questions and/or implementation of algorithms covered in the lectures by the instructor and guest talks. Collaboration is permitted, but the assignments must be submitted independently. You need to list the names of collaborators (if any).

Class Presentation:
In each class, there will be one or two 25-minute student presentations of the research papers. Some classes will be guest talks and lectures by the instructor. Each presentation will be followed by a discussion of the paper. You must sign up for a presentation at least one week in advance. Each student will have to present 1-2 times during the semester. Presentations will be graded based on technical understanding of the material, clarity, and organization.

Class Participation: You will be graded based on attendance, and participation in class discussions during and after the presentation.

Grading Scheme:

45%	Homework
35%	In-class Quizzes
10%	Class Presentation(s)
10%	Class Participation

Tentative Course Calendar (subject to change; see updates on the course website)

Date	Day	Seminar Topic	Homework
August			
26	Mon.	Logistics and Overview of the Topics	
28	Wed.	Probability and Linear Algebra Review	HW1 Release
September			
2	Mon.	Labor Day; No Class	
4	Wed.	Queueing Intro; Scheduling for Parallel Computing	
9	Mon.	Grid Computing; MapReduce	
11	Wed.	Tail at Scale; Straggler Replication	HW1 Due; HW2 Release
16	Mon.	Sparrow; Attack of the Clones	
18	Wed.	Task Replication in Queuing Systems	
23	Mon.	Coding Theory and Codes for Storage	
25	Wed.	Speeding up ML using Codes; Rateless Codes	HW2 Due; HW3 Release
30	Mon.	**Quiz 1**	
October			
2	Wed.	Learning a Code; Gradient Coding	

7	Mon.	Machine Learning & SGD Introduction	
9	Wed.	Survey of SGD methods; Convergence Analysis	
14	Mon.	DistBelief; ImageNet Classification	HW3 Due; HW4 Release
16	Wed.	HogWild; Slow and Stale Gradients	
21	Mon.	PipeDream; Stale Synchronous Parallel	
23	Wed.	**Quiz 2**	
28	Mon.	Elastic Averaging SGD; Cooperative SGD	
30	Wed.	AdaComm; Federated Learning	HW4 Due; HW5 Release
November			
4	Mon.	Fed Prox; Multi-task Learning	
6	Wed.	TernGrad; Model Compression	
11	Mon.	ATOMO; PowerSGD	
13	Wed.	MATCHA; Fed Learning with non-IID data	HW5 Due; HW6 Release
18	Mon	MAB Intro; Bayesian Opt Intro	
20	Wed.	Thanksgiving break; No class	
25	Mon.	HyperBand; Neural Architecture Search	
27	Wed.	Spearmint; Parallel Bayesian Opt	
December			
2	Mon.	Final Class	
4	Wed.	**Quiz 3**	HW6 Due

Education Objectives (Relationship of Course to Program Outcomes):
(a) Know the state-of-the-art frameworks used in cloud and machine learning infrastructure, and the algorithms that make them work
(b) Provide constructive criticism of other's research, and identify open research directions
(c) Collaborate in multi-disciplinary teams and identify, formulate, and solve engineering problems
(d) Effectively organize and present research ideas

ECE Academic Integrity Policy
(http://www.ece.cmu.edu/programs-admissions/masters/academic-integrity.html**):**

The Department of Electrical and Computer Engineering adheres to the academic integrity policies set forth by Carnegie Mellon University and by the College of Engineering. ECE students should review fully and carefully Carnegie Mellon University's policies regarding Cheating and Plagiarism; Undergraduate Academic Discipline; and Graduate Academic Discipline. ECE graduate student should further review the Penalties for Graduate Student Academic Integrity Violations in CIT outlined in the CIT Policy on Graduate Student Academic Integrity Violations. In addition to the above university and college-level policies, it is ECE's policy that an ECE graduate student may not drop a course in which a disciplinary action is assessed or pending without the course instructor's explicit approval. Further, an ECE course instructor may set his/her own course-specific academic integrity policies that do not conflict with

university and college-level policies; course-specific policies should be made available to the students in writing in the first week of class.

This policy applies, in all respects, to this course.

CMU Academic Integrity Policy (http://www.cmu.edu/academic-integrity/index.html**):**

In the midst of self-exploration, the high demands of a challenging academic environment can create situations where some students have difficulty exercising good judgment. Academic challenges can provide many opportunities for high standards to evolve if students actively reflect on these challenges and if the community supports discussions to aid in this process. It is the responsibility of the entire community to establish and maintain the integrity of our university.

This site is offered as a comprehensive and accessible resource compiling and organizing the multitude of information pertaining to academic integrity that is available from across the university. These pages include practical information concerning policies, protocols and best practices as well as articulations of the institutional values from which the policies and protocols grew. The Carnegie Mellon Code, while not formally an honor code, serves as the foundation of these values and frames the expectations of our community with regard to personal integrity.

This policy applies, in all respects, to this course.

The Carnegie Mellon Code

Students at Carnegie Mellon, because they are members of an academic community dedicated to the achievement of excellence, are expected to meet the highest standards of personal, ethical and moral conduct possible.

These standards require personal integrity, a commitment to honesty without compromise, as well as truth without equivocation and a willingness to place the good of the community above the good of the self. Obligations once undertaken must be met, commitments kept.

As members of the Carnegie Mellon community, individuals are expected to uphold the standards of the community in addition to holding others accountable for said standards. It is rare that the life of a student in an academic community can be so private that it will not affect the community as a whole or that the above standards do not apply.

The discovery, advancement and communication of knowledge are not possible without a commitment to these standards. Creativity cannot exist without acknowledgment of the creativity of others. New knowledge cannot be developed without credit for prior knowledge. Without the ability to trust that these principles will be observed, an academic community cannot exist.

The commitment of its faculty, staff and students to these standards contributes to the high respect in which the Carnegie Mellon degree is held. Students must not destroy that respect by their failure to meet these standards. Students who cannot meet them should voluntarily withdraw from the university.

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Cheating
(http://www.cmu.edu/academic-integrity/cheating/index.html) states the following:

According to the University Policy on Academic Integrity, cheating "occurs when a student avails her/himself of an unfair or disallowed advantage which includes but is not limited to:

- Theft of or unauthorized access to an exam, answer key or other graded work from previous course offerings.
- Use of an alternate, stand-in or proxy during an examination.
- Copying from the examination or work of another person or source.
- Submission or use of falsified data.
- Using false statements to obtain additional time or other accommodation.
- Falsification of academic credentials."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Plagiarism
(http://www.cmu.edu/academic-integrity/plagiarism/index.html) states the following:

According to the University Policy on Academic Integrity, plagiarism "is defined as the use of work or concepts contributed by other individuals without proper attribution or citation. Unique ideas or materials taken from another source for either written or oral use must be fully acknowledged in academic work to be graded. Examples of sources expected to be referenced include but are not limited to:

- Text, either written or spoken, quoted directly or paraphrased.
- Graphic elements.
- Passages of music, existing either as sound or as notation.
- Mathematical proofs.
- Scientific data.
- Concepts or material derived from the work, published or unpublished, of another person."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Unauthorized Assistance
(http://www.cmu.edu/academic-integrity/collaboration/index.html) states the following:

According to the University Policy on Academic Integrity, unauthorized assistance "refers to the use of sources of support that have not been specifically authorized in this policy statement or by the course instructor(s) in the completion of academic work to be graded. Such sources of support may include but are not limited to advice or help provided by another individual, published or unpublished written sources, and electronic sources. Examples of unauthorized assistance include but are not limited to:

- Collaboration on any assignment beyond the standards authorized by this policy statement and the course instructor(s).
- Submission of work completed or edited in whole or in part by another person.
- Supplying or communicating unauthorized information or materials, including graded work and answer keys from previous course offerings, in any way to another student.
- Use of unauthorized information or materials, including graded work and answer keys from previous course offerings.
- Use of unauthorized devices.
- Submission for credit of previously completed graded work in a second course without first obtaining permission from the instructor(s) of the second course. In the case of concurrent courses, permission to submit the same work for credit in two courses must be obtained from the instructors of both courses."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Research Misconduct
(http://www.cmu.edu/academic-integrity/research/index.html) states the following:

According to the University Policy For Handling Alleged Misconduct In Research, "Carnegie Mellon University is responsible for the integrity of research conducted at the university. As a community of scholars, in which truth and integrity are fundamental, the university must establish procedures for the investigation of allegations of misconduct of research with due care to protect the rights of those accused, those making the allegations, and the university. Furthermore, federal regulations require the university to have explicit procedures for addressing incidents in which there are allegations of misconduct in research."

The policy goes on to note that "misconduct means:

- fabrication, falsification, plagiarism, or other serious deviation from accepted practices in proposing, carrying out, or reporting results from research;

- material failure to comply with Federal requirements for the protection of researchers, human subjects, or the public or for ensuring the welfare of laboratory animals; or

- failure to meet other material legal requirements governing research."

"To be deemed misconduct for the purposes of this policy, a 'material failure to comply with Federal requirements' or a 'failure to meet other material legal requirements' must be intentional or grossly negligent."

To become familiar with the expectations around the responsible conduct of research, please review the guidelines for Research Ethics published by the Office of Research Integrity and Compliance.

This policy applies, in all respects, to this course.

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you have questions about this or your coursework, please let me know.