Electrical & Computer ENGINEERING

Carnegie Mellon

Course Syllabus

18-847F: Special Topics in Computer Systems
Foundations of Cloud and Machine Learning Infrastructure
Fall 2017

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

Number of Units: 12
Pre-requisites: None

Instructor: Prof. Gauri Joshi
Email ID: gaurij@andrew.cmu.edu
Office Location: CIC 4105
Office Hours: Wed 2:00 pm - 3:00pm, or by appointment

Course Description: The objective of this graduate-level seminar course is to introduce students to cloud and machine learning infrastructure, and its theoretical foundations. 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. For the final group project you will write a short research paper on a topic of your choice.

The first half of the course will cover distributed computing and storage systems. We will study frameworks such as MapReduce and Spark, 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 stochastic gradient descent, hyper-parameter tuning, and deep reinforcement learning.

Course Website: http://andrew.cmu.edu/~gaurij/18-847F-Fall-2017.html.The latest course calendar, along with the list of upcoming speakers will be posted here

Course Canvas: http://www.cmu.edu/canvas. You should use the Canvas site for submitting assignments and projects, and viewing grades and instructor feedback.

Learning Objectives
(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

Suggested Reference Textbooks:

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.

Homework:
Students will be divided into two groups. Each group will read one research paper. By 9:00 am on the day of each class, each student must submit a review of the paper. Collaboration is permitted, but the reviews must be written independently. You need to list the names of collaborators (if any). The review consists of a summary of the strengths and weaknesses of the paper, and some questions to facilitate discussion in class.

Class Presentation:
In each class, one student from each group will give a 20-minute presentation of the research paper they read before class. Each presentation will be followed by a discussion of the strengths and weaknesses of the paper and possible future directions. You must sign up for a presentation at least one week in advance. Each student will have to present 2-3 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.

Final Project:
You will form groups of 2-3 and complete a project which can be original research, or an in-depth literature survey, or implementation of a theoretical concept. Drawing the project topics from your own graduate research is permitted and encouraged. At the end of the semester you will submit a report in the form of an ACM conference-style paper (maximum 5 pages), and present it in class.

Grading Scheme:

45%	Homework
20%	Presentations
10%	Class Participation
25%	Final Project

Tentative Course Calendar

Date	Day	Seminar Topic
August		
28	Mon.	No Class
30	Wed.	Intro and Logistics
September		
4	Mon.	Labor Day; No Class
6	Wed.	Overview of Cloud and ML Infrastructure
11	Mon.	Scheduling for Parallel Computing, Warehouse Computing
13	Wed.	Tail at Scale, MapReduce
18	Mon.	Spark, Sparrow
20	Wed.	Attack of the Clones, Straggler Replication
25	Mon.	Task Replication in Queueing Systems
27	Wed.	Guest Lecture: Cloud Spot Markets
October		
2	Mon.	RAID Systems, Coding for Distributed Storage
4	Wed.	Locality and Repair in Distributed Storage (Dimakis, Gopalan)
9	Mon.	Guest Lecture: Hitchhiker codes, EC-Cache; Project Proposals Due
11	Wed.	Coded MapReduce, Speeding up Machine Learning
16	Mon.	Stochastic Grad. Descent (SGD) and Support Vector Machines
18	Wed.	Survey of SGD methods
23	Mon.	Guest Lecture: Short-Dot Coded Computation
25	Wed.	Backpropagation, LeNet
30	Mon.	AlexNet, GoogleNet
November		
1	Wed.	Distributed Deep Learning (Dean), TensorFlow
6	Mon.	Mini-batch SGD and its convergence (Shamir, Bottou)
8	Wed.	SVRG methods, Model-Runtime Trade-offs in Distributed SGD
13	Mon.	Asynchronous SGD Analysis, YellowFin
15	Wed.	Hyper parameter tuning
20	Mon	Generative Adversarial Networks
22	Wed.	Thanksgiving break; No class
27	Mon.	Deep Reinforcement Learning; Draft of project report due
29	Wed.	Project Presentations
December		
4	Mon.	Project Presentations
6	Wed.	Project Presentations
8	Fri.	Final Project reports due

Course Wiki:
Students are encouraged to use the ECE wiki to provide feedback about the course at: http://wiki.ece.cmu.edu/index.php.

Academic Integrity Policy:
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 and CIT's policies regarding Cheating and Plagiarism.

Students at Carnegie Mellon are engaged in preparation for professional activity of the highest standards. In any presentation, creative, artistic, or research, it is the ethical responsibility of each student to identify the conceptual sources of the work submitted. Failure to do so is dishonest and is the basis for a charge of cheating or plagiarism, which is subject to disciplinary action.

Cheating includes but is not necessarily limited to:

1. Submission of work that is not the student's own for assignments or exams.
2. Submission or use of falsified data.
3. Submission of the same work for credit in two courses without obtaining the permission of the instructors beforehand.

This policy applies, in all respects, to 18.847F.