Title: Syllabus Slug: syllabus Date: 2019-12-20

Welcome to AC295 (http://harvard-iacs.github.io/2020F-AC295). In this course we explore advanced practical data science practices. The course will be divided into three major topics:

1. How to scale a model from a prototype (often in jupyter notebooks) to the cloud. In this module, we cover virtual environments, containers, and virtual machines before learning about microservices and Kubernetes. Along the way, students will be exposed to Dask.

2. How to use existing models for transfer learning. Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks. This can be very important, given the vast compute and time resources required to develop neural network models on these problems and given the huge jumps in skill that these models can provide to related problems. In this part of the course we will examine various pre-existing models and techniques in transfer learning.

3. In the third part we will be introducing a number of intuitive visualization tools for investigating properties and diagnosing issues of models. We will be demonstrating a number of visualization tools ranging from the well established (like saliency maps) to recent ones that have appeared in https://distill.pub.

Lectures: Tuesday and **Thursday** 10:30-11:45am
TFs:
Office Hours: TBD

List of Contents

- Course Policies

Prerequisites

You are expected to be fluent in programming (Python), statistics knowledge at the level of Stat 110 or above, data science (or machine learning) at the level of AC209A and AC209B.

Software

We will be using a variety of software primarily Python 3, Pytorch, Tensorflow, Docker, etc. More details in class.

Topics

The course is organized in three modules.

1. **Deploy data science** (integration + scalability)
   1a. Virtual Environments, Virtual Boxes, and Containers
   1b. Kubernetes
   1c. Dask

2. **Transfer learning and distillation**
   2a. Intro to Transfer Learning: basics and Convolutional Neural Networks (CNNs) review 2b. Transfer Learning for Images and SOTA Models 2c. Language Models and Transfer Learning with Text Data 2d. Transformers and Attention 2e. Distillation and Compression

3. **Visualization as investigative tool**
   3a. Introduction and Overview of Viz for Deep Models: lime and shapley 3b. CNNs for Image Data, Activation Maximization and Saliency Maps 3c. Attention for Debugging Language Models

Course Activities

Each module is structured in three types of activities and they are: **Lectures**, **Reading Discussion**, and **Practicum**. Each activity requires the students to complete different assignments in the form of exercise/homework, discussions, reading assignment, and presentation (see Assignments below). During the

first weeks of each module, students will attend Lecture on Tuesday and Reading Discussion on Thursday. The last week of each module will be Practicum. **Attendance is mandatory**.

1. **Reading List** consists in papers, blogs and other reading material that will be released no later than the beginning of each week. This will be the base for all the activities during the week See Readings Guidelines here link to guidelines.
2. **Lectures** are held online **Tuesdays** from 10:30-11:45am (and possibly depedning on timezone of students repeat **Tuesday** from 9:00-10:15pm). During this activity we will discuss and summarize the basic concepts of the material covered during the week.
3. **Journal Discussions** are held online on **Thursdays** from 10:30-11:45am. During this activity, two groups will present to the rest of the class one or two papers from the Reading List and lead the discussion. See Paper Presentation Guidelines here link to guidelines.
4. **Practicum** are activities in the form of a project based on the material covered in the module. The students will work in groups and be expected to deliver a complete assignment in 10 days. There will be two practicums.

Assignments

The **final grade** will be calculated using the following weights for each assignment:

Exercises

There are 8 exercises to complete. They will be released at the end of each regular week Lecture and due the next one. The exercises are graded on a scale 1 to 5, where 5 is the highest grade.

Discussion Forum

There will be a discussion forum the day before the Reading Discussion based on what the reading content. The question will cover some of the material from Reading List and students will access them using Ed Platform on Canvas (select Ed from tab on course main page).

Presentations

At every Reading Discussion, groups will present the reading material assigned at the beginning of the week. Please see these on the presentations.

Practicums

There will be two practimum during the first two modules (see schedule for details).

Final Projects

There will be a final group project due during Exams period encompassing all the material learned in class.

Assignment	Final Grade Weight
Discussion Forum	10%
Exercises	10%
Presentations	15%
Practicums	45%
Final Projects	20%

|**Total**| **100%**|

Getting Help

For questions about exercise, course content, package installation, and after you have tried to troubleshoot yourselves, the process to get help is:

1. Go to **Office Hours**, this is the best way to get help.
2. Post the question in **Ed Forum** and hopefully your peers will answer.

Course Policies

Collaboration Policy

We encourage you to talk and discuss the assignments with your fellow students. Discussion is encouraged. Presentation during Reading Discussion, Practicum and Projects are group activities.

Communication from Staff to Students

Class announcements will be through **Ed Forum**.

Academic Honesty

Ethical behavior is an important trait of a Data Scientist, from ethically handling data to attribution of code and work of others. Thus, in AC295 we give a strong emphasis to Academic Honesty. As a student your best guidelines are to be reasonable and fair. We encourage teamwork for problem sets, but you should not split the assignments and you should work on all the problems together.

Accommodations for students with disabilities

Students needing academic adjustments or accommodations because of a documented disability must present their Faculty Letter from the Accessible Education Office(AEO) and speak with Pavlos by the end of the third week of the term: Friday, September 18. Failure to do so may result in us being unable to respond in a timely manner. All discussions will remain confidential.