Course Information:

Note: Starting Monday January 31, we are meeting in Maxwell Dworkin G115

Instructor: Michael P. Brenner

Teaching Fellows: Liyuan Chen, Megan Engel, Javin Pombra and Shiva Mudide

Class Schedule: Monday and Wednesday 9-1015am EST; Friday 9-1015am will contain the section for the class. We will attempt to have all class sessions, including sections, recorded.

Piazza: link in canvas and also here

Section Times: Friday 9-1015

Office Hours: (See Zoom section for the zoom links to these office hours)

Javin: Wednesday 8-10pm

Michael: Tuesday 9-10pm

Megan: Thursday 12-2pm

Liyuan: Saturday 2-4pm

Shiva: Saturday 6:30-8:30pm

E-mail: brenner@seas.harvard.edu

Course Overview:

The scientific endeavor is about solving inverse problems. Given data, our task is to invent a model that describes the data. Until recently, the predominant paradigm for solving this problem in the physical sciences was find a physical or mechanistic model of the phenomenon in question, and find the variant of the model that explained the data. This would be done by understanding in detail the features in the data that a class of models would predict, and then fitting these features to the data. The fits would allow narrowing down the classes of the model. A second paradigm dominated those cases where causal mechanisms were unknown â€“ here the goal is to simply try to find a function that captures the correlation structure of the data.

Recent advances in computation and data analysis have changed things significantly. Given sufficient data, it is now possible for computers to learn sophisticated features themselves using neural networks and large scale computation. In a range of practical examples, this has proven to surpass human ability to engineer features. This advance has had tremendous practical applications (think Amazon/Apple/Google/etc.) but it also promises to make a significant impact on science and engineering practice. There seems to be enormous opportunity to change the way that science is done â€“ but exactly how to go about using these methods to make science progress is not clear.

The goal of this class is think about these broad questions in a pedagogical fashion. This is on one hand going to be a research class, in that we are primarily motivated by understanding how current advances can improve scientific practice. On the other hand, we will do this in a pedagogical fashion: We will review classical methods for solving inverse problems â€“ ranging from Bayesian parametric methods, to those based on the laws of physics and in this context, and will also introduce and study neural network based approaches. Theory will be mixed with computation and practical examples.

Course Structure:

The semester itself will be divided into the following three parts:

1. In the first part, we will discuss pedagogical material, focusing on both the mathematical foundations as well as software packages that have been recently developed to help implement these ideas.

2. The second part of the semester will focus on applications. In each class we will discuss recent progress in some area of applying the ideas of this class to scientific problems. In section, you will implement models yourselves using python notebooks. During this second part of the semester there will not be weekly problem sets but instead you will work in small groups on mini projects. We will define a number of projects that are related to the applications we discuss as a class and you will work through them in small groups.
3. In the last part of the semester, class will shift to focusing on final projects. How this is structured will depend on the enrollment, but the idea is for you to define yourself what you will work on.

Class (required): There will be two lectures per week.

Class will consist of the following two components:

- Lectures, which will introduce various forms of inverse problems and paradigms for solving them. Lectures will cover the mathematical and theoretical aspects of these problems, as well as introduce the computational tools we will be using
- Guest speakers, who will be brought in to discuss their own papers on the cutting edge of the literature

Section (required): Computation and computational exercises are at the heart of this class and as such we will have a mandatory section each week. In these sections, you will walk through a jupyter notebook to learn how you can use the computational tools introduced in lecture to actually solve inverse problems. Sections will be recorded and made available on Canvas.

Problem Sets and Projects (required): For the first ~2/3 of the class, there will be weekly problem sets which you will do in the form of a jupyter notebook. The main problem in every problem set will ask you to solve an inverse problem â€“ we will give you data and your task will be to figure out how we generated it. We will run Kaggle competitions each week to evaluate how you and your peers did. The winner of the competition for each problem set will be invited to present their solution to the class. **You will not be graded on your Kaggle ranking**, it is simply a fun way to work on solving inverse problems as a class. The last ~1/3 of the class will not have any problem sets, you will instead complete a mini-project and a final project during this time.

Tutorials (optional): Optional tutorials will be offered on an as-needed basis to help students get up to speed on the computational components of the course. Currently, we are planning to offer a tutorial on Python and object oriented programming, as well as a tutorial on deep learning with TensorFlow 2. Additional tutorials may be added depending on student demand. These tutorials will be recorded and made available on Canvas.

Supplementary Materials: We will give out supplementary material (readings, video, code, etc.) each week to help you better understand the topics covered in lecture and section. These will primarily be drawn from the book *Probabilistic Machine Learning: An Introduction* by Kevin Murphy, as well as jupyter notebooks from the books associated GitHub. This book is not yet published, but the draft that we will be using can be found here. The GitHub repository can be found here. We will also be including supplementary material from various other resources throughout the semester, you can find all the materials for each week in the weekly overviews under the modules tab. The supplementary materials will be organized into three categories:

- Required: These materials will be required reading/watching. We will try to keep these materials as concise as possible to maximize their usefulness.
- Recommended: These materials are recommended in order to help you with the problem sets each week.
- Optional: These materials are completely optional, and intended for students who have time and want to dive deeper.

Grading/Requirements:

Homework. (40 %) For the first âˆ¼ 2/3 of the class, there will be regular homework, that will mainly involve computational implementations of the ideas of the class. Coding will be done in Python, where there are easy to use high level libraries for implementing the ideas of the class (for example, Keras for Neural networks.) The main problem in every problem set will ask you to solve an inverse problem â€“ we will give you data and your task will be to figure out how we generated it. We will run Kaggle competitions each week to evaluate the work. **You will not be graded on your Kaggle ranking**, it is simply a fun way to work on solving inverse problems as a class.

Final projects: (30 %, including class presentations) A key part of the class is for each student to carry out a final project, applying the ideas we are discussing to some problem of particular interest. To prepare for a final project, students will make presentations on papers working up to the project throughout the class. Final projects can be carried out in groups of up to 3 people. We will start to choose them and work on them after about the first month of the class.

Mini-projects (30 %) During the second part of the class when we are discussing applications, instead of homework you will work in small groups on mini projects. These projects will be defined by the course staff around the application areas we are discussing.

Additional Information

A detailed schedule of the course is given here: Detailed Schedule

A shorter PDF syllabus is here: syllabus_2022.pdf