New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 9223 - I
Visualization for Machine Learning

Instruction Mode: Blended (Online & In-Person)

Thursday 2:00PM-4:30PM
2 Metro Tech Center Rm 8.012

Spring 2021

Course Instructor

Cláudio T. Silva
Tandon: 370 Jay Street, Brooklyn, room 1153, 646 997-4093
Center for Data Science: 60 5th Ave, Manhattan, room 701, 212-998-3091
csilva@nyu.edu

Teaching Fellows

Gromit Yeuk-Yin Chan
370 Jay Street, 11th floor
gromit.chan@nyu.edu

Jorge Piazentin Ono
370 Jay Street, 11th floor
jorgehpo@nyu.edu

COVID 19

The instruction mode for this class is **blended**, what means that students can attend the class in person or remotely. Students are encouraged to participate in class synchronously, but all sessions will be recorded.

Seat tracking is mandatory, and students will select a seat during the first day of classes.

All office hours will use Zoom. Discord will be the main platform for instructor / student communications. **You will find details on how to log in to the class Discord on NYU classes.** We will also use GitHub, and details will be forthcoming.

If you are experiencing an illness or any other situation that might affect your academic performance in a class, please email Deanna Rayment, Coordinator of Student Advocacy, Compliance and Student Affairs: deanna.rayment@nyu.edu. Deanna can reach out to your instructors on your behalf when warranted.

Course Pre-requisites

You should have solid programming expertise, of the level expected from a first-year graduate student in computer science.

We expect that you have a solid foundation in **either** data visualization or machine learning. But ultimately, your background will determine the type of project that you work on. If you are only well-versed in machine learning, then you will need to learn the basics behind data visualization. On the other hand, if you are only a data visualization expert, then you will need to learn the basic components of machine learning.

Course Description

This course is a research-oriented course on topics related to visualization for machine learning. Our course is based on foundations of visual analytics, which is an area of data visualization that is concerned with improving a human's analytic process, or how one makes sense of data for a given problem: understanding, reasoning, and making decisions about a provided dataset, and a given problem domain. Visual analytics, in particular, is concerned with combining automated processes with human-driven processes that are built around data visualization: visual representations of data, and ways to interact with data. Given the rapid growth in machine learning in the last decade, research in visual analytics has witnessed similar growth in leveraging machine learning in a variety of ways.

Acknowledgement

This course is based on the Visual Analytics and Machine Learning course designed by **Professor Matthew Berger** (Vanderbilt University). We first offered it at Tandon on Spring 2020.

For this second offering, we are updating the content of the course, in particular with more practical material aimed at enabling students to experience data analysis tasks through visualization.

We are also making modifications to adapt the content and delivery to the unique situation of the COVID-19 pandemic**.**

Course Objectives

This course is designed to sharpen a student's knowledge of visualization and machine learning, and how the two areas interact. It is expected that the student will be a more effective data scientist by being fluent on the connections between the two areas. It is also designed around a major project, which will help the student develop research skills.

Course Structure

The course will include lectures, labs, and student presentations. We will strive to have hands-on sessions to complement theoretical materials.

The course starts with a short primer on visualization. We will introduce machine learning concepts as they are needed in the class. After covering the use of synthetic data for benchmarking machine learning techniques, we will cover while-box and black-box machine learning explainers. We will continue with dimensionality reduction (clustering) techniques (e.g., t-SNE). AutoML will be covered next, followed by ways to measure and assess ML fairness. Topology Data Analysis will be covered next, followed by multiple lectures on visualizing deep neural networks.

Readings

There is no textbook for the course - most lectures will be based on recent technical papers, which have not yet been incorporated into textbooks. We will have suggested reading materials for each class. It is expected that, prior to the lecture, you have read the corresponding papers.

Here are supplemental readings to be used as reference material:

1. Machine Learning: a Probabilistic Perspective, Kevin Patrick Murphy, MIT Press, 2012.
2. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, Christoph Molnar, https://christophm.github.io/interpretable-ml-book/
3. Visualization Analysis and Design, Tamara Munzner, A K Peters Visualization Series. CRC Press, 2014.
4. Deep Learning, Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press, 2016 http://www.deeplearningbook.org/

Research Project

The bulk of the course will be devoted to a semester-long research project. Please see the project section of the course for more details. As part of the project, you will be expected to reproduce prior work, as well as implement a proposed research idea of your choosing (selected in consultation with the instructor). Moreover, you will be expected to demonstrate both the prior work, and your final research project, to the class during lectures. Again, please see project for additional details. **Projects are expected to be pursued in groups of two. Once the group is finalized, students cannot change or separate their groups throughout the semester.**

Course Assessment

- Assignments (20%)
- Project Proposal: 15%
    - Presentation: 5%
    - Proposal Document: 10%
- Related Work: 20%
    - Presentation and Demonstration: 10%
    - Source Code and Documentation: 10%
- Project Updates: 5% (total)
- Full Project: 35%
    - Presentation: 10%
    - Full Submission: 25%
- Class Participation: 5%

Course Schedule

The course schedule is tentative and will be adjusted along the way.

Lecture 1: Introduction to Visualization for Machine Learning – Part I

Lecture 2: Introduction to Visualization for Machine Learning – Part II
- Assignment #1 out

Lecture 3: Synthetic Data & Benchmarks

Lecture 4: AutoML
- Assignment #2 out

Lecture 5: White Box Algorithms

Lecture 6: Black Box Algorithms
- student presentation (3 slots of 10 minutes)

Lecture 7: Dimensionality Reduction
- student presentation (3 slots of 10 minutes)

Lecture 8: Student Project Proposals

Lecture 9: Fairness
- student presentation (3 slots of 10 minutes)

Lecture 10: Topology Data Analysis
- student presentation (3 slots of 10 minutes)

Lecture 11: Deep Learning Model Understanding – Part I
- student presentation (3 slots of 10 minutes)

Lecture 12: Deep Learning Model Understanding – Part II
- student presentation (3 slots of 10 minutes)

Lecture 13: Deep Learning Model Understanding: GANs
- student presentation (3 slots of 10 minutes)

Lecture 14: Project Presentations

Moses Center Statement of Disability

If you are a student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

NYU School of Engineering Policies and Procedures on Academic Misconduct

– Complete Student Code of Conduct can be found here.

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
3. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
4. Unauthorized collaboration: working together on work meant to be done individually.
5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

NYU School of Engineering Policies and Procedures on Excused Absences

– Complete policy can be found here.

A. Introduction: An absence can be excused if you have missed no more than 10 days of school. If an illness or special circumstance has caused you to miss more than two Lectures of school, please refer to the section labeled Medical Leave of Absence.

B. Students may request special accommodations for an absence to be excused in the following cases:

1. Medical reasons
2. Death in immediate family
3. Personal qualified emergencies (documentation must be provided)
4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the Coordinator of Student Advocacy, Compliance and Student Affairs and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

NYU School of Engineering Academic Calendar

– Official calendar can be found here.

This course does not have a final exam.

Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu