PROJECTS IN ADVANCED MACHINE LEARNING

Columbia University
GR5074, Spring 2022
Thurs, 6:10PM-8:00PM (EST)
330 URIS HALL

Instructor:	Marco Morales
Email:	marco.morales@columbia.edu
Office:	509E International Affairs Building
Office Hours:	TBD, and by appointment

TA:	Rachel Lee	Heinrich Peters
Email:	yl3751@tc.columbia.edu	heinrichpeterz@gmail.com

I. Overview

Machine learning algorithms continue to advance in their capacity to predict outcomes and rival human judgment in a variety of settings. This course is designed to offer insight into advanced machine learning models, including Deep Learning, Convolutional Neural Networks for image and text data, Object detection models, Recurrent Neural Networks (time-series data), and Adversarial Neural Networks.

Roughly half of the course will engage machine learning methods while the other half of the course will be devoted to students working in key substantive areas, where advanced machine learning will prove helpful — like computer vision and images, text and natural language processing, and tabular data. Students will be tasked to develop team projects in these areas and they will develop a public portfolio of three (or four) meaningful projects.By the end of the course, students will be able to show their work by launching their models in live REST APIs and web-applications.

Prerequisites: basic probability and statistics, basic linear algebra and calculus. Students are also expected to have familiarity with using **Python**, the `scikit-learn` package, and GitHub.

II. Course Resources

Textbooks: There are no required textbooks for this course, but you will find these to be very useful in addition to the lectures and course readings:

- François Chollet. *Deep Learning with Python.* Manning Publications, Greenwich, CT, first edition, 2017 [DLWP]
- Andreas C. Müller and Sarah Guido. *Introduction to Machine Learning with Python.* O'Reilly Media, Boston, MA, 2016 [IMLP]
- Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow.* O'Reilly Media, Boston, MA, first edition, 2017 [HOML]
- Max Kuhn and Kjell Johnson. *Applied Predictive Modelling.* Springer, New York, NY, 2013 [APM]
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer, New York, NY, second edition, 2009 [ESL]
- Ian Goodfellow, Joshua Bangio, and Aaron Courvill. *Deep Learning.* The MIT Press, Cambridge, MA, 2016 [DL]

Additional materials: other articles and materials will be distributed, which cover additional topics in more depth.

III. Course Dynamics

Attendance and Class Participation. Your attendance and participation are necessary at every meeting. We will put applied machine learning models into practice via labs and model development competitions throughout the course, so a laptop is required.

Project Assignments. Students will have two (2) *Homework Project Assignments* to complete throughout the semester, as well as one *Final Project Assignment* on a topic of their choosing. These Project Assignments will fit into the overall Data Science workflow. Specific instructions, format and deadlines will be given as the semester progresses.

Late Submission Policy: All class assignments are expected to be submitted on the due date. Please note that 10% of the maximum grade will be deducted from the score for every day the assignment is submitted late.

IV. Course requirements

The grade for this course will depend on the fulfillment of three main requirements:

(i) Homework Project Assignments (60%)

(ii) Final Project Assignment (25%)

(iii) Attendace and Participation (15%)

V. Course Outline

Class	Topic	Readings/Notes
1	Intro to course	
2	Python, Github workflow setup, Colab, review of supervised learning / review tabular data preprocessing / Practice on multiple datasets	IMLP Ch 1, Intro to column transformer, DLWP Ch. 1
3	Project #1 - Tabular Data Mini-hackathon Work together as a team to build models individually and improve them collectively. Compare modelling approaches. Choose best model and launch it into a live REST API and web-application.	
4	Neural Networks; Convolutional neural networks for image classification (Assignment #1 Due)	IMLP p104-109, DL Ch 6, Ch 7.8; DL Ch 7.12, Ch 9, keras docs
5	Advanced Architectures; Transfer Learning; Wrangling image data	DLWP Ch. 5, DL Ch 9, HOML, Ch 13
6	Project #2 – Image Data Mini-hackathon Work together as a team to build models individually and improve them collectively. Compare modelling approaches. Choose best model and launch it into a live REST API and web-application.	
7	Object detection in image data, YOLO and other architectures (Assignment #2 Due)	Yali Amit and Pedro Felzenszwalb, Object Detection
8	Intro to Recurrent Neural Networks (Sequence/Time-series data)	HOLM, Ch 14, DLWP Ch. 6 Intro, 6.2-3
9	RNN (i.e. LSTM) data exploration and Model Dev Examples	
10	Text/NLP data preprocessing and model generation	IMLP p323-336
11	LSTM (Recurrent Neural Nets) for text classification	DLWP Ch. 6.1,6.4-6.5
12	Project #3 – Text Data Mini-hackathon Work together as a team to build models individually and improve them collectively. Compare modelling approaches. Choose best model and launch it into a live REST API and web-application.	HOLM, Ch 14
13	Intro to Adversarial neural networks (Assignment #3 Due)	DLWP Ch. 8.5,Goodfellow et al. Gen. Adversarial Nets

Statement on Academic Integrity

Columbia's intellectual community relies on academic integrity and responsibility as the cornerstone of its work. Graduate students are expected to exhibit the highest level of personal and academic honesty as they engage in scholarly discourse and research. In practical terms, you must be responsible for the full and accurate attribution of the ideas of others in all of your research papers and projects; you must be honest when taking your examinations; you must always submit your own work and not that of another student, scholar, or internet source. Graduate students are responsible for knowing and correctly utilizing referencing and bibliographical guidelines. When in doubt, consult your professor. Citation and plagiarism-prevention resources can be found at the GSAS page on Academic Integrity and Responsible Conduct of Research.

Failure to observe these rules of conduct will have serious academic consequences, up to and including dismissal from the university. If a faculty member suspects a breach of academic honesty, appropriate investigative and disciplinary action will be taken following the Dean's Discipline procedures.

Statement on Disability Accommodations

If you have been certified by Disability Services (DS) to receive accommodations, please either bring your accommodation letter from DS to your professor's office hours to confirm your accommodation needs, or ask your liaison in GSAS to consult with your professor. If you believe that you may have a disability that requires accommodation, please contact **Disability Services** at 212-854-2388 or disability@columbia.edu.

Important: To request and receive an accommodation you must be certified by DS.