COMS 4721: Machine Learning for Data Science

Course Information
1/12/2021

Prof. John Paisley

Columbia University

COURSE INFORMATION

Time: T/TH 11:40am-12:55pm on Zoom via Courseworks (live & recorded)
Instructor: Prof. John Paisley
Email: jwp2128@columbia.edu
Office Hour: Wednesday 12pm-1pm on Zoom

Course website:
http://www.columbia.edu/~jwp2128/Teaching/COMS4721.html
Suggested readings will be put on the website.
Lecture slides will be uploaded to Courseworks before class.

Teaching Assistants:
1. Erica Wei (uni: cw3137), OH: W 8-10pm @ Zoom
2. Wei Zhang (uni: wz2363), OH: F 7-9pm @ Zoom
3. Arunesh Mittal (uni: am4589), OH: W 6-8pm @ Zoom
4. Aditya Makkar (uni: am5284), OH: M 10pm-12am @ Zoom

Office hours will begin the second week. Information will be posted under the
Announcements tab on Courseworks.

GRADING

4 homeworks (15% each)

- A mix of written questions and coding.
- Your answers (in PDF file) and code to be uploaded to Courseworks.

final exam (40%)

- Take home at end of semester, date to be determined.
- Will cover entire course.

The class will be curved. The vast majority will get a B or higher.

CODING & RESOURCES

- ▶ Any coding language is acceptable for the homeworks. Data will be provided in .csv files.

- ▶ This is not a class for learning how to code. The TA's and I cannot help you with coding questions.

- ▶ I will set up a Piazza forum where you can post your questions to the class. Feel free to use the forum for other questions besides coding tips.

- ▶ In all homework assignments you must turn in original source code. (Python and R notebooks are not considered source code.)

- ▶ The TA's and I will have office hours. We won't monitor Piazza and we won't monitor any questions on Canvas. Please email me directly.

ACADEMIC HONESTY

Please see:
http://www.cs.columbia.edu/education/honesty

On the homework: Can consult each other to get/give help on the problems, but write-up and code must be your own.

"Help" that results in significantly overlapping written answers or code will be considered a violation by all involved (see below).

During the tests: No notes of any kind can be used.

Violators: Lose a lot of points.

HOMEWORK POLICIES (APPEARS ON EACH HOMEWORK)

Please read these instructions to ensure you receive full credit on your homework. Submit the written portion of your homework as a *single* PDF file through Courseworks (less than 5MB). In addition to your PDF write-up, submit all code written by you in their original extensions through Courseworks (e.g., .m, .r, .py, .c). Any coding language is acceptable, but do not submit notebooks, do not wrap your files in .rar, .zip, .tar and do not submit your write-up in .doc or other file type. When resubmitting homeworks, be sure to resubmit *all files*. Your grade will be based on the contents of one PDF file and the original source code. Additional files will be ignored. We will not run your code, so everything you are asked to show should be put in the PDF file. Show all work for full credit.

Late submission policy: Late homeworks will have 0.1% deducted from the final grade for each minute late. *Your homework submission time will be based on the time of your __last__ submission to Courseworks. I will not revert to an earlier submission!* Therefore, do not re-submit after midnight on the due date unless you are confident the new submission is significantly better to overcompensate for the points lost. Submission time is non-negotiable and will be based on the time you submitted your last file to Courseworks. The number of points deducted will be rounded to the nearest integer.

TEXTBOOKS

I will suggest (overlapping) readings from the following two books. The sections will be posted on the website and are just meant to be pointers. You may find resources you like better, which is ok (e.g., Wikipedia).

- ► T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning, Second Edition*, Springer. (see link on website for book)

- ► C. Bishop, *Pattern Recognition and Machine Learning*, Springer. (see link on website for supplemental materials)

- ► H. Daume, *A Course in Machine Learning*, Draft. (see link on website)

RESTRICTIONS AND PREREQUISITES

This class is not open to those who have taken an equivalent course at Columbia, e.g., ELEN 4720, COMS 4771, STATS 4240 & 4400 or IEOR 4525.

There are no course prerequisites for the class. However, I will assume you know the basics of and feel somewhat comfortable with:

- ▶ calculus
- ▶ linear algebra
- ▶ probability and statistical concepts
- ▶ coding and data manipulation