

COMS 4721: Machine Learning for Data Science

Course Information

1/17/2017

Prof. John Paisley

Columbia University

COURSE INFORMATION

Instructor: Prof. John Paisley

Email: jpaisley@columbia.edu

Office Hour: Monday 11-12, Mudd 422

Course Website:

<http://www.columbia.edu/~jwp2128/Teaching/W4721/Spring2017/W4721Spring2017.html>

Suggested readings will be put on the website.

Lecture slides will be uploaded to Courseworks before class.

Teaching Assistants:

1. Ghazal Fazelnia, gf2293@columbia.edu
2. Tianhao Lu, tl2710@columbia.edu
3. Dheeraj Kalmekolan, drk2143@columbia.edu
4. Ashutosh Nanda, an2655@columbia.edu
5. Francisco Javier Arceo, fja2114@columbia.edu
6. TBD

GRADING

5 homeworks (12% each)

- A mix of written questions and coding.
- PDF and code to be uploaded to Courseworks.

mid-term exam (20%)

- Written, in class on March 7 (before spring break).
- Will cover first half of course.

final in-class exam (20%)

- Written, in class on April 27 (last day of class).
- Not comprehensive. Will cover second half of course.

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

No R credit or Pass/Fail options.

CODING

- ▶ Any coding language is acceptable for the homeworks. Data will be provided in .csv text files.
- ▶ This is not a class for learning how to code. The TA's and I cannot help you with coding questions (but feel free to ask your classmates).
- ▶ 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.

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 identical 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: Will get a zero on the homework/test in question.

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, etc.). Any coding language is acceptable. Do not wrap your files in .rar, .zip, .tar and do not submit your write-up in .doc or other file type. 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

Each class will have (highly overlapping) readings from the following two books. These 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., COMS 4771, STATS 4240, STATS 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!