September 1, 2020

AMSC 808N / CMSC 828V

Numerical Methods

for

Data Science

and

Machine Learning

Maria Cameron

Department of Mathematics

Department of Computer Science (affiliate)

What is Data Science?

- Wiki: is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.
- Briefly: Data Science is a science about learning from data.
- Birth: J. Tukey's paper "The Future of Data Analysis" (1962)

Father-founder:
John Tukey (1915—2000)
(best known for the FFT algorithm and box-plot)

"50 Years of Data Science" (2015)

Activities of data Science

1. Data Exploration and Preparation;
2. Data Representation and Transformation;
3. Computing with Data;
4. Data Modeling;
5. Data Visualization and Presentation;
6. Science about Data Science.

David Donoho
Anne T. And Robert M. Bass
Professor of Humanities and Sciences
Professor of statistics

Successes of DS&ML that struck me personally

- ML for solving PDEs in high dimensions:

https://arxiv.org/pdf/1710.00211.pdf,

https://arxiv.org/pdf/1802.10275.pdf

https://arxiv.org/pdf/1906.06285.pdf

- ML for geophysics:

https://seg.org/Education/Lectures/Distinguished-Lectures/2020-DL-Fomel

- ML for molecular dynamics:

https://www.sciencedirect.com/science/article/pii/S0959440X19301551

Inspiration for this course

David Bindel

Associate Professor of Computer Science

Director, Center for Applied Math

CS, applied math, math, civil engineering, and CSE, and data science.
Confusing rabbits since 2003.

425 Gates Hall
Dept of Computer Science
Cornell University
Ithaca, NY 14853-5169

OH: By appointment
Scheduler link
bindel@cs.cornell.edu

Research highlights

Optimizing stellarators
Advancing magnetic confinement fusion through optimization and hidden symmetries.

Home page Announcement

Nonlinear waves in resonant MEMS

Teaching highlights

Numerical Methods for Data Science
Revisiting numerical methods teaching for modern data science applications.

UChicago (June 2019) SJTU (May 2019) UMD (April 2019)
SJTU (June 2018) Cornell (Spring 2018)

Background

- Linear algebra: vectors, matrices, maps, norms, inner products
- Calculus: $f(x+p) = f(x) + \nabla f(x)^T p + (1/2) p^T \nabla^2 f(x) p + \ldots$
- CS: notation $O(n^p)$, operation count, sparse vs dense
- Matlab: basic syntax
- Floating point arithmetic: machine epsilon, roundoff
- Conditioning: ill- vs well-conditioned, condition number

See Bindel's note:
http://www.cs.cornell.edu/~bindel/class/sjtu-summer18/lec/background.pdf

Logistics

- Lectures Tuesday and Thursday 9:30 — 10:45 AM via zoom
- Homework: latex it, upload your pdf file on ELMS (30% of your grade)
- Programming projects:
  - Working groups (2-3 students)
  - Any suitable language, I recommend Matlab or Python
  - Each group member should develop his/her code, then verify results and write a common report with other group members; link all codes to it; upload pdf on ELMS.
  - Peer review: this should not be time-consuming
  - Finally, I check peer-reviewed reports
- Final exam: asynchronous, due in one week after posting

Asking questions and getting answers

- Piazza
- Office hours starting Sept. 9: via zoom
  - Monday 1:30—2:30 PM
  - Wednesday 4:30—5:30 PM
- Email: `mariakc@umd.edu`

Request to you:

- Please type in the zoom chat:
  - Your name
  - Your email
  - Your program and your year in it
  - Have you taken AMSC660? AMSC661?