CS205L

Continuous Mathematical Methods
(emphasizing Machine Learning)

http://web.stanford.edu/class/cs205l/index.html

CS 205 Course History

- Mathematical Modeling of Continuous Systems (Tomasi)
  - ‘94-’95, ‘95-’96, ‘96-’97, ‘97-’98, ‘98-’99, ‘99-2000, 2000-’01 (7 years)
  - 2001-’02 not taught
- New Emphasis on Robotics, Vision & Graphics (Fedkiw)
  - CS205: 2002-’03, ‘03-04, ‘04-’05, ‘05-’06, ‘06-’07 (5 years)
  - CS205A: 2007-’08, ‘08-’09, ‘09-’10, ‘10-’11, ‘11-’12, ‘12-’13 (6 years)
  - Justin Solomon (@MIT now) 2013-’14, ‘14-’15 (2 years)
  - Doug James 2015-’16, ‘16-’17, ‘17-’18 (3 years)
- New Emphasis on Machine Learning (Fedkiw)
  - CS205L: 2018-’19, ‘19-’20, ‘20-’21, ‘21-’22, ’22-’23, ‘23-’24 (6th year)

GPU History

- 1970s Specialized hardware for graphics became popular for Arcade games and home consoles
- 1993 Nvida was founded (Jensen Huang et al.)
- 1994 The term “GPU” was coined by SONY, referring to a chip in the 1994 Playstation 1
- 1995 Toy Story (Pixar) revolutionized 3D graphics for animated feature films
- 1996 Super Mario 64 (Nintendo 64) revolutionized 3D graphics for games
- 1999 Star Wars: Episode 1, The Phantom Menace revolutionized 3D movie VFX
- 2002 restructured CS205 to focus on graphics applications

- Now (like graphics in the past) machine learning is the popular computationally intensive area of computer science that is getting its own specialized hardware (e.g. Nvidia’s ML/DL chips)
- So, restructured CS205 (once again) to CS205L

CS205L: Continuous Mathematical Methods with an Emphasis on Machine Learning

A survey of numerical approaches to the continuous mathematics used throughout computer science with an emphasis on machine and deep learning. Although motivated from the standpoint of machine learning, the course will focus on the underlying mathematical methods including computational linear algebra and optimization, as well as special topics such as automatic differentiation via backward propagation, momentum methods from ordinary differential equations, CNNs, RNNs, etc. Written homework assignments and (straightforward) quizzes focus on various concepts.

Replaces CS205A, and satisfies all similar requirements.

Prerequisites: Math 51; Math 104 or 113 or equivalent or comfort with the associated material.

Lectures

- Slides are broken into units, and each unit is broken into smaller chunks (via different background colors)
- There is separate slide deck on notation (see the web site)
- A lot of the connections between continuous mathematics and machine/deep learning is still cutting-edge; so, Ask Questions!

- Lectures are recorded, and attendance is optional
- Problem sets are based on the lectures and (related) supplemental readings
- Quizzes are based on the lectures (and the questions are handed out in advance)

Problem Sets (50% of grade)

- Written (no programming)
- Assigned weekly on Thurs, covers Tues/Thurs lectures of **that** week (due 1 week later)
- Up to 4 people in a discussion group (list names of group on HW)
- However, each person turns in their own unique-as-possible write-ups
  - if your write-up is not original enough, you may be asked to do extra problems for that assignment
- Each problem is assigned to a specific CA/CAs, and CAs only address their problem
  - However, CAs will address all *general* course material (and quiz questions)
- Each problem is graded (coarsely) by the relevant CA/CAs: 0, ¼, ½, ¾, 1 points
- Problems vary in difficulty: 1, 2, 3 stars
- Choose 6 stars worth of problems (each week, the maximum score is 6 points)

Weekly Quizzes (50% of grade)

- Each Thurs, potential quiz questions will be released (with the problem set)
- Quiz questions cover both lectures (Tues/Thurs) of that week
- You may discuss questions with your HW discussion group, the CAs, etc.
- The following Monday afternoon (4 days later), you will take a one-on-one (no partners) oral quiz with one of the CAs (online)
- They will choose 1 question from the pre-determined list for you to answer
- If you need special accommodations for the time slot, reach out to the CAs

Roadmap

- Part I – Linear Algebra (units 1-12) $Ac = b$
- Part II – Optimization (units 13-20)
  - (units 13-16) Optimization -> Nonlinear Equations -> 1D roots/minima
  - (units 17-18) Computing/Avoiding Derivatives
  - (unit 19) Hack 1.0: "I give up" $H = I$ and J is mostly 0 (descent methods)
  - (unit 20) Hack 2.0: "It's an ODE!?" (adaptive learning rate and momentum)

Theory

Methods

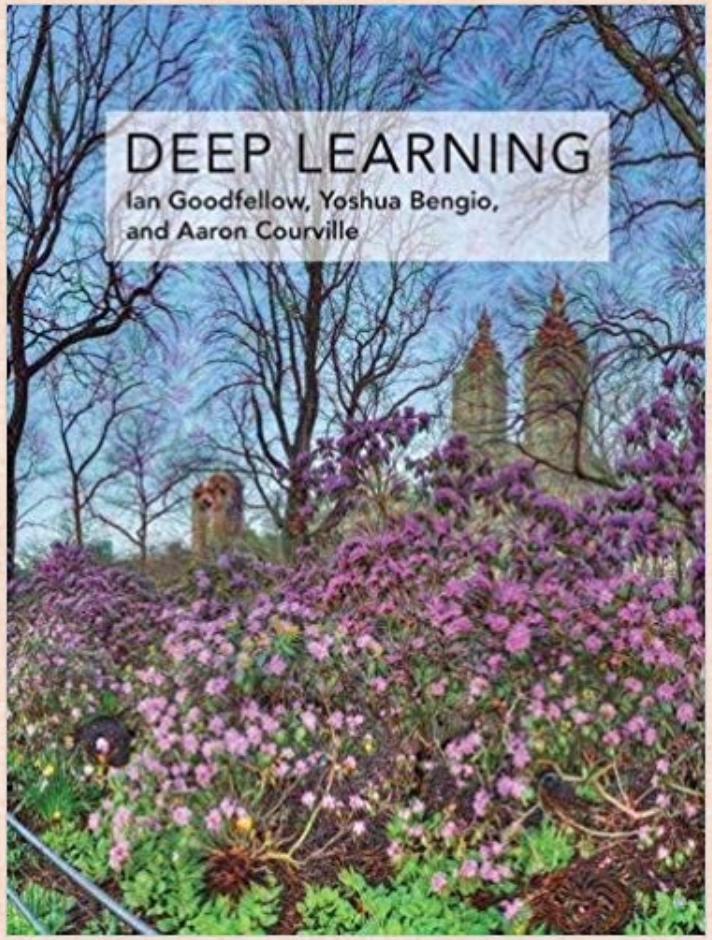
Solving Linear Systems

- Theory, all matrices: SVD (units 3, 9, 11)
- Square, full rank, dense:
  - LU factorization with pivoting (unit 2)
  - Symmetric: Cholesky factorization (unit 4), Symmetric approximation (unit 4)
- Square, full rank, sparse (iterative solvers) (unit 5):
  - SPD (sometimes SPSD): Conjugate Gradients
  - Nonsymmetric/Indefinite: GMRES, MINRES, BiCGSTAB (not steepest descent)
- Tall, full rank (least squares to minimize residual) (unit 8):
  - normal equations (units 9, 10), QR, Gram-Schmidt, Householder (unit 10)
- Any size/rank (minimum norm solution) (unit 11):
  - Pseudo-Inverse, PCA approximation, Power Method (unit 11)
  - Levenberg-Marquardt (iteration too), Column Space Geometric Approach (unit 12)

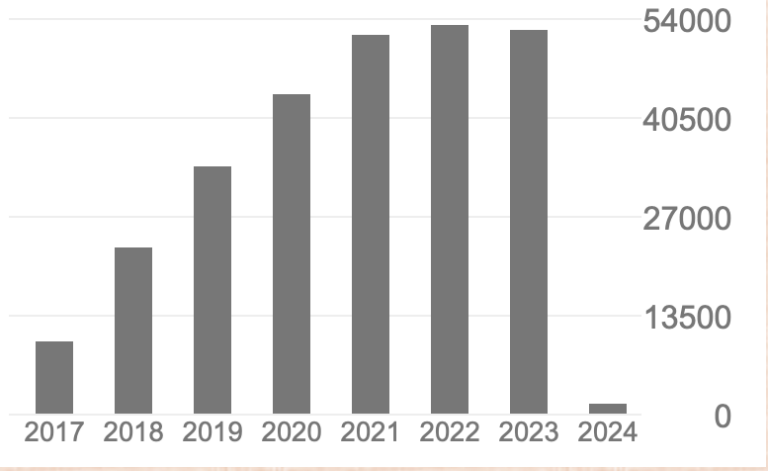
-----Original Message-----

From: Ian Goodfellow

Ron, I don't know if you remember me, but I took your class circa 2009. Writing numerically stable code was a big factor in my success as a DL researcher, especially for GANs.



	All	Since 2019
Citations	276439	236549
h-index	90	84
i10-index	144	139



Questions?