Course Logistics

Optimization and Computational Linear Algebra for Data Science

Marylou Gabrié

The teaching team Website

- **Lecturer**: Marylou Gabrié – *mgabrie at nyu.edu*

  Office Hours: Mondays 3-4pm (Zoom) + by appointment

- **Sections leaders**:

Colin Wan	Ying Wang	Zahra Kadkhodaie
Tuesdays 8am	Tuesdays 9am	Tuesdays 4pm
O.H.: Wednesday 2-3pm	O.H.: Tuesdays 10-11am	O.H.: Tuesdays 5-6pm

- **Website**
  `marylou-gabrie.github.io/linalg-for-ds.html`

Course components

Three main components:

1. Lectures

   Introduces the concepts

2. Recitations

   Practice!

3. Homeworks

   Helps you master concepts and methods

Course components

Three main components:

1. Lectures

   Introduces the concepts

2. Recitations

   Practice!

3. Homeworks

   Helps you master concepts and methods

- Feedback, remarks about the lectures / recitations / homeworks ... :
  - email me!
  - link for anonymous feedback on the course's website.

Course components

Three main components:

1. Lectures

   Introduces the concepts

2. Recitations

   Practice!

3. Homeworks

   Helps you master concepts and methods

- Feedback, remarks about the lectures / recitations / homeworks ... :
  - email me!
  - link for anonymous feedback on the course's website.

Grades:

1. Weekly homeworks (40%)
2. Exams: Midterm (25%) + Final (35%)

Weekly timeline

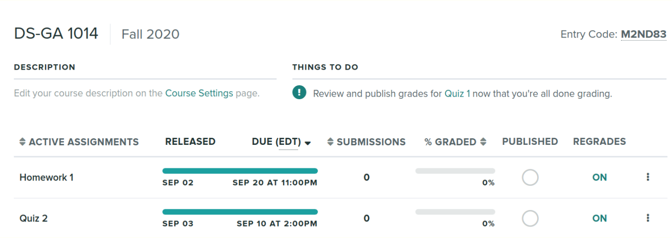
September 2021

‹ Today ›

Sun	Mon	Tue	Wed	Thu	Fri	Sat
29	30	31	Sep 1	2 **Session 1** September 2, 2021 Class	3	4
5	6	7 **Labs 1** September 7, 2021 Labs	8	9 **Session 2** September 9, 2021 Class	10	11
12 **HW1 due** September 12, 2021 HW	13	14 **Labs 2** September 14, 2021 Labs	15	16 **Session 3** September 16, 2021 Class	17	18

Homeworks

- Homeworks questions are available on the **course's webpage** and have to be submitted on **Gradescope**.



DS-GA 1014 | Fall 2020

Entry Code: M2ND83

DESCRIPTION

Edit your course description on the Course Settings page.

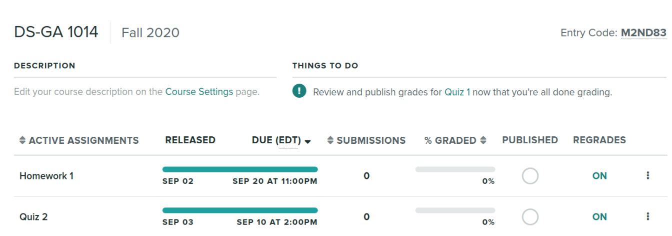
THINGS TO DO

Review and publish grades for Quiz 1 now that you're all done grading.

ACTIVE ASSIGNMENTS	RELEASED	DUE (EDT)	SUBMISSIONS	% GRADED	PUBLISHED	REGRADES
Homework 1	SEP 02	SEP 20 AT 11:00PM	0	0%		ON
Quiz 2	SEP 03	SEP 10 AT 2:00PM	0	0%		ON

Homeworks

- Homeworks questions are available on the **course's webpage** and have to be submitted on **Gradescope**.



- We will not accept late homeworks.

- I encourage you to type your homeworks using LaTeX.

  Some instructions and template available on the course's webpage.

- Otherwise, you can scan your handwritten work. **It has to be legible!!!**

Midterm and Final

- **Midterm** ($\sim$ mid-October) and **Final** will be «in-person exams» (unless something changes).
- Limited time and open book (notes are ok / search the web is not ok)

Check out the syllabus on the course webpage!

Questions on logistics ?

Contents

1. Linear algebra
   About 2/3 of the lectures
2. Convex optimization
   About 1/3 of the lectures
3. Overview of the lectures
   A quick look at the menu

Linear algebra

Why linear algebra?

« Linear algebra $\simeq$ geometry in arbitrary dimension »

Why do we need to do geometry ?

- In many case, our data is a collection of « data points » that are points $(x_1, \dots, x_n)$
- To understand the structure of our data, we have to investigate the geometry of our data points: are they divided into clusters? are they «aligned» ?
- When $n = 1, 2, 3$, one can easily plot our data, but what about $n = 10000$?

Applications

You will learn linear algebra, while studying applications for data science such as:

- Data compression
- Principal component analysis

    Find directions along which the variance of the data is maximal

- Dimensionality reduction

    Reduce the dimension of a dataset while preserving its structure

- Linear regression
- Google's Page Rank and Markov chains

    Ranking any objects that can be compared!

Optimization

Optimization

In machine learning, we often have to minimize functions

$$f(\theta) = \mathrm{Loss}(\mathrm{data}, \mathrm{model}_\theta) \quad \text{with respect to} \quad \theta \in \mathbb{R}^n.$$

- For $n = 1, 2$, one could plot f to find the minimizer.
- This is intractable for larger dimension.

We will

- focus on convex cost functions f.
- study gradient descent algorithms to minimize f.

Overview of the lectures

Outline

1. Vectors and vector spaces
2. Linear transformations and matrices
3. The rank
4. Norm and inner product
5. Eigenvalues, eigenvectors and Markov chains
6. The spectral theorem and PCA
7. Graphs and Linear Algebra
8. Convex functions
9. Linear regression
10. Optimality conditions
11. Gradient descent