Hubert Wagner
College of Liberal Arts & Sciences

College *of* Liberal Arts *and* Sciences

Home

MTG 7396/MAT 4930, Computational Applied Topology course, Fall 24

MAS 4115, Linear Algebra for Data Science, Spring 2025

MAS 4115, Linear Algebra for Data Science, Spring 2025

Instructor: Hubert Wagner, hwagner[…]ufl.edu

Time and Location:

Section 0059 | M,W,F | Period 6 (12:50 PM – 1:40 PM), Location: LIT 0233

Section 8993 | M,W,F | Period 7 (1:55 PM – 2:45 PM), Location: LIT 0233

Office Hours: M,W,F | Period 8 (3:00 PM – 3:50 PM), Location: my office LIT 428 (or zoom on request).

General Course Description: A second course in linear algebra, focusing on topics that are the most essential for data science. Introduces theory and numerical methods required for linear problems associated with large data-sets and machine learning.

Specific to our sections:

Fair Warning: It is not obvious from the general course description, but in the sections taught by me emphasis is on practical, hands-on problem-solving in the context of data-science. In other words: **there will be a lot of non-trivial python programming, data analysis and discussions**. See below for more details.

Another warning: This is **not an easy class** – be prepared to put a significant effort. In the past, students who did not put an effort (and for example tried – and failed – to use ChatGPT to solve homework assignments) had to drop the class or failed the course.

Our Focus: We will highlight linear algebra concepts using practical examples in data analysis, many of which coming from my experience in industry. We will learn the programming tools necessary to handle such data, with focus on image data (with some sound and text). Also, much emphasis will be on developing intuitions (often through geometric visualization) and communication skills necessary for working in data science.

Course Goals and Objectives: A student who successfully completes this course will be able to:

Map data analysis problems to concepts of high-dimensional linear algebra.

Think critically about the choice of the method and the result it produces.

Use Python and its libraries (mostly numpy, matplotlib, scipy, sklearn, keras) to solve concrete data analysis problem.

Articulately discuss and clearly explain mathematical concepts in the context of data science.

Continue learning more advanced techniques on their own or by taking more specialized courses.

Schedule overview:

Week 1-2: Review of (very) basic Linear Algebra: vectors, Euclidean norm and distance. Representing data as a high-dimensional vector space. Basic data analysis techniques and problems based on these concepts. In parallel: intro to Python + numpy and matplotlib.

Week 3-4: Nonlinear dimensionality reduction (focusing on tSNE) and clustering (k-means). Basic concepts related to optimization, gradient descent. More Python and libraries, handling image data.

Week 5-6: More Linear Algebra: dot product, orthogonality; techniques based on hyperplanes (SVMs, kd-trees); Elements of supervised learning, with focus on classification tasks.

Week 7-10: More advanced Linear Algebra: matrices, linear transformations, matrix decompositions focusing on SVD; Problems and techniques based on these concepts (including linear dimensionality reduction, Johnson-Lindenstrauss lemma, low-rank approximations, Moore-Penrose pseudoinverse, linear regression).

Week 11-14: Neural networks (MLP, CNNs, GANs), backpropagation, various loss functions, elements of information theory (entropy, cross entropy, relative entropy). Intro to the Keras library.

Week 15: Summary.

Logistics:

Prerequisites: A course in linear algebra (MAS 3114, MAS4105 or equivalent course) and calculus (MAC 2313) are required.

Programming Prerequisite: Class demos, examples and homework assignments will **use Python**. However, you **are not expected** to be proficient in Python at the start of the course. You **are expected** to have enough programming experience in another language to pick up the basics of Python quickly.

The first couple of weeks of classes are meant to help you pick up Python and its

basic libraries (numpy, matplotlib, scipy, later sklearn and Keras). The initial homework assignments will be assigned smaller weights so that early mistakes will not have a decisive effect on the final grade.

We will use google colab (https://colab.research.google.com/) and similar online programming environments, so no complicated setup of a programming environment is necessary. This requires you to have a google account (it's free).

If you have a laptop, please bring it to each class. (If you don't we'll find you a partner with a laptop.)

Participation: This is a synchronous, face-to-face class.

Work and grading:

The grade will depend on:

> homework: 60%
>
> exams: 30%
>
> participation and activity: 10%

Homework. Homework will be posted on Canvas and you will upload your solution in the Assignments section before the stated deadline. Usually you will have one week to complete the homework, which will be graded within a couple of days (from a deadline). **All homework assignments will be mini-projects and will require programming.** You can expect six assignments of varying size and difficulty.

Exams: There will be two or three exams in total (including the final exam). The format will be announced for each exam. The exam tasks will check your understanding of details and of the theoretical foundations of the used methods. They will not require programming.

Activity and participation: I will encourage discussion during classes. Good questions, answers, comments will be rewarded. We will have a discord server for online discussion, constructive activity there will also be rewarded.

Grading. The grade ranges for the total scores will be no tougher than: 93-100% A; 90-92% A-; 88-89% B+; 83-87% B; 80-82% B-; 78-79% C+; 73-77% C; 70-72% C-; 60-69% D; <60% F.

The outlined arrangements may change based on University guidelines and student needs. We will discuss and finalize them during the first week of classes.

Additional information:

Resources: The course will be based on lecture notes and programming notebooks on colab. The following resources may be useful as additional references, but are not required:

Linear Algebra and Learning from Data, by Gilbert Strang, Wellesley-Cambridge Press; First edition (2019).

Mathematical Foundations for Data Analysis, by Jeff M. Philips, online version: Springer, 2021 (online version, 95% complete: https://mathfordata.github.io/versions/M4D-v0.6.pdf

Neural Networks and Deep Learning by Michael Nielsen http://neuralnetworksanddeeplearning.com/index.html

Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron Courville http://www.deeplearningbook.org/

Honor Code and Collaboration: In this course authorized aid on projects and hw consists of talking to me, other students, reading the documentation for your computational platform, and looking at the notes for this course. This means that you are not allowed to look on-line, in other books specifically for solutions to the hw or projects, or at the written solutions of other students. Looking up general stuff like definitions, usage of Python libraries is of course fine. You can collaborate with fellow students but must write up and code individually.

Excused Absences: In certain circumstances a student will be able to make up a missed exam. These circumstances could include medical situations, family emergencies, travel for University activities (eg. band, debating club, etc), and religious observances. In these cases the student must inform me before or within one week after the missed work and **provide written documentation**.

Additional Information:

Grades: Grading will be in accord with the UF policy stated at https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx.

Honor Code: "UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class."

Class Attendance: "Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at: https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx."

Grading Disputes: Any issues or questions about the grading of homework or exams

must be brought to my attention within one week after the exams or homework are returned to the class.

Diversity Statement: I am committed to diversity and inclusion of all students in this course. I acknowledge, respect, and value the diverse nature, background and perspective of students and believe that it furthers academic achievements. It is my intent to present materials and activities that are respectful of diversity: race, color, creed, gender, gender identity, sexual orientation, age, religious status, national origin, ethnicity, disability, socioeconomic status, and any other distinguishing qualities.

Accommodations for Students with Disabilities: "Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester."

Online Evaluations: "Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/."

Contact information for the Counseling and Wellness Center: https://counseling.ufl.edu/, 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

U Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit umatter.ufl.edu/ to refer or report a concern and a team member will reach out to the student in distress.

UF | UNIVERSITY *of* FLORIDA
The Foundation for The Gator Nation