MAS 4115: Linear Algebra for Data Science Spring 2025

Instructor:	Youngmin Park
E-mail:	park.y@ufl.edu
Office:	454 Little Hall
Office Hours:	LIT 454, WF 11:45am–12:35pm (per. 5)
Class Meetings:	LIT 233, MWF 10:40am–11:30am (per. 4)
Prerequisites:	MAS3114, MAS4105 or equivalent

Course Description and Objectives: 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. Topics include general data manipulation in Python, k-means clustering, linear classifiers, linear and non-linear dimension reduction methods (SVD and tSNE), gradient descent, and artificial neural networks (e.g., multilayer perceptrons and convolutional neural networks). To prevent cheating with Chat GPT and similar tools, there will be a significant in-class component to this course (see "Grading" below).

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

- Explain data science concepts at low and high levels.
- Perform basic linear algebra computations in Python.
- Understand which data science methods are appropriate for which types of data.
- Construct and train various data science models.
- Construct and train simple feedforward neural networks.

Course Materials: All lecture materials (Colab notebooks, hand-written notes) will be posted to Canvas. The following may be useful supplementary resources but are not necessary for this course:

- Linear Algebra and Learning from Data, by Gilbert Strang, Wellesley-Cambridge Press; First edition (2019).
- Numerical Linear Algebra, Lloyd Trefethen and David Bau, SIAM Press, 1997
- 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/

Programming Prerequisite: We will use Python. You will need to have enough experience with a programming language to pick up Python reasonably quickly. We will use Keras along with standard numerical and plotting packages (Numpy, Scipy, and Matplotlib).

Grading: All grades will be posted to Canvas.
Any issues or questions about the grading of homework assignments or quizzes must be brought to my attention within one week after the assignment or quiz grades are posted to canvas.

- **Homework** (50%): Homework will be assigned every other week on Fridays. The lowest homework score will be dropped.
  - I will not accept late assignments. If the submission is closed on Canvas, do not contact me.
  - There will be audits of homework assignments for 20–50% of the class per assignment. We will sit down in person and I will ask you questions about your code.
  - Use of Google Colab is **mandatory** because Colab notebooks contain a full revision history. Do not write code elsewhere and copy/paste into the notebook. Save frequently and incrementally. Failure to follow this direction will result in a zero for the assignment.
- **Quizzes** (20%): There will be weekly in-class quizzes on Wednesdays. I am not a fan of in-class evaluations, but this is an unfortunate necessity with the advent of tools like ChatGPT.
- **In-class Participation** (20%): I will select random people to answer questions (see **In-Class Participation** below).
- **Attendance** (10%): Attendance will be taken at the start of each class.

Semester letter grade assignments will be no stricter than the following: 93-100 A, 90-92 A-, 87-89 B+, 83-86 B, 80-82 B-, 77-79 C+, 73-76 C, 70-72 C-, 67-69 D+, 63-66 D, 60-62 D-, 0-59 E. We will adhere to the university grading policies that can be found here: https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/

Weekly Schedule (subject to change):

- Week 1-2: Review of (very) basic Linear Algebra: vectors, Euclidean norm and distance. Data as a high-dimensional vector space. Basic data analysis techniques and problems based on these concepts (including similarity search, k-means). Intro to Python + numpy.
- Week 3-4: Basic concepts related to optimization, gradient descent; nonlinear dimensionality reduction (focusing on tSNE). More Python and libraries.
- Week 5-6: More Linear Algebra: dot product, orthogonality; techniques based on hyperplanes (SVMs, kd-trees).
- Week 7-10: More advanced Linear Algebra: matrices, linear transformations, various matrix decompositions focusing on SVD; Problems and techniques based on these concepts (including linear regression, (linear) dimensionality reduction, low-rank approximations).
- Week 11-14: Various neural networks, backpropagation, various loss functions, elements of information theory. Intro to Keras library.

- Week 15: Summary.

Announcements: You are responsible for all announcements made in Canvas and via email which could include changes in exam dates and material covered.

Attendance: Attendance is required. We will adhere to the university attendance policies that can be found here: `https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/`.

In-Class Participation:
In-class participation is a necessary evil.
- As questions are asked in class, students will be selected at random.
  - Students may ask for clarification.
  - Students may "pass" on the question for a zero.
  - I will work with the selected student towards an answer, but there will be no partial credit.
- We will sometimes perform in-class programming exercises. You may be expected to share your screen via Zoom.

Etiquette, Expectations, Assumptions:
- It is the students' responsibility to start the homework early. I do not care about loss of internet access or family deaths. Two weeks is enough time to account for potential issues that may come up near the submission deadline.
- It is the students' responsibility to understand the instructor. I will not cater to your preferences.
- Some students might find the instructions vague or incomplete. It is your job to figure out what I mean. Students will not be spoon-fed every detail.
- Students will follow instructions **thoroughly** and **thoughtfully**. This will take time and effort on the part of the student. Instructions will not contain every single detailed step needed to complete the assignment.
  Example: If the instructor repeatedly says in class that vectorization is important, students are expected to vectorize in the assignment even if the instructions do not explicitly say so. Moreover, it is expected that the student will do the work to figure out whether or not something can be vectorized.
- Students are expected to double check that their solution(s) fully address the question.
- Students will do their due diligence to solve problems prior to contacting the professor. Due diligence includes searching error and warning messages prior to contacting the instructor.
  Example: while using plt.imshow, you see the warning, "Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers)". The image does not render as expected and instead shows a blank white image.
  To debug the problem, copy/paste the warning into google or a chatbot and see what others have said about the message. This is a good example where the student is expected to debug the problem on their own.
  Non-Example: After receiving an error, do not just say "it didn't work." or "I got an error." Look at what the error says.
- Just because the instructor did not list something here does not mean it is fair game to pretend like it was never implied. This syllabus is not a literal checklist. Again, it is up to you to make sure that you have not missed something that I consider obvious.
  Example: You are expected to use the knowledge/skills/methods learned in class. If I see that you've solved a problem using skills or methods from elsewhere, you will earn a zero.
  Example: We will use SVD on 2D arrays. Using SVD on 3D arrays and solving an assignment using these 3D SVDs (and additional oddities like subtracting the mean) will warrant a zero.
- I will be more forgiving with mistakes in the second half of the semester PROVIDED THAT students follow instructions carefully and make an earnest effort.
- Do not use ChatGPT or similar chatbots to mindlessly answer questions. It's obvious when zero thought was put into an assignment, whether or not the code is perfect. Save yourself some embarrassment and don't bother submitting.

Asking for Hints and Using Chat GPT (or similar chatbots like Gemini, Copilot, Perplexity):
- Always cite your sources.
  - If you use Chat GPT (or a similar tool) on an assignment, you must include the link to the conversation with your submission.
  - If you work with classmate(s) on an assignment, list their names in the submission.
  - **Failing to cite sources is equivalent to plagiarism and may warrant a failing score.**

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.

Honor Code and Collaboration: In this course authorized aid on projects and homework 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 may not look up solutions to assignments or quizzes online, in other books, or copy from other students (but using these resources to gain a better understanding of the material is fair game). You can and should

collaborate with fellow students, but you must write code individually.

Student 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/`.

Disabilities statement: 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/`. 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.

Academic Integrity: 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 Conduct Code specifies a number of behaviors that are in violation of this code and the possible sanctions. See `https://sccr.dso.ufl.edu/process/student-conduct-code/` to read the Conduct Code. If you have any questions or concerns, please consult with the instructor.