MAS 4115: Linear Algebra for Data Science Spring 2024

Instructor:	Youngmin Park
E-mail:	park.y@ufl.edu
Office:	454 Little Hall
Office Hours:	TBD, or appointment
Class Meetings:	MWF 12:50pm–1:40pm (per. 6), MAT 6
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 QR and singular-value decompositions of matrices; conditioning and stability; complex vector spaces; the DFT and linear filters; deep learning; fully connected and convolutional nets; and gradient descent.

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

- Perform basic linear algebra computations by hand and in Python.
- Prove the existence of the various standard matrix decompositions and use their numerical implementation for data analysis and solving linear problems.
- Construct routines which avoid the common sources of error based on an appreciation of conditioning and stability in numerical linear algebra computations.
- Derive the basic properties and write Python code implementations of the Discrete Fourier Transform, convolution, and filtering
- Construct simple feedforward neural networks using learning functions, loss functions and stochastic gradient descent

Course Materials: The course will be based on lecture notes (which will be provided via Google Colab with links posted in Canvas). The following may be useful supplementary resources:

- 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: Class demos will use Python. Class assignments will require Python or comparable platform, such as R, Matlab or Julia. Otherwise, you will need to have enough experience with a programming language to pick up Python reasonably quickly. In particular, we will use a significant amount of Keras along with standard numerical and plotting packages (Numpy, Scipy, and Matplotlib).

Grading: All homework grades and quiz grades will be posted to canvas. **Please notify me as soon as possible if you anticipate being unable to submit a homework assignment**. I cannot help if you come to me at the end of the semester to tell me that you were sick and could not submit an assignment.
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** (80%): Homework will be assigned every other week on Fridays and due the in two weeks with occasional breaks. It will be posted as a file in Canvas. The homework will foster mastery over the material covered in class in the previous weeks. It will include hand computations, proofs and computer computations. The lowest homework score will be dropped.
- **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.

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/

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 on-line, 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 must write up and code individually.

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.

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.

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/`.

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 or TA in this class.