Lecture 0: Introduction

Lecturer: Jie Fu, Ph.D.

EEL 3850

What you'll learn

Foundations of data science:

- Introduction to probability and statistics.
- Introduction to applied linear algebra.

Practical data science engineering applications:

- Learn Python - the most popular programming language and for Data Science and Software Development.
- Demonstrate proficiency in using Python libraries such as Pandas & Numpy, and developing code using Jupyter Notebooks.
- Access and web scrape data using APIs and Python libraries.

Meeting time and place

Attending lectures are required.

- Period 5 (11:45 AM - 12:35 PM) MWF Lar320 and Period 9 (4:05 PM - 4:55 PM) Th. CSE-E119

Exams:

- **Exam 1: Feb 10 (part 1) and Feb 11(part 2)**

- **Exam 2: March 13 (part 1) and March 14 (part 2)**

- **Exam 3: April 22 (part 1) and April 23 (part 2) [Different from one.ufl]**

Teaching Team

- **Instructor: Jie Fu** (Assistant Professor, ECE)
- Office: MALA 4106
- Email: fujie@ufl.edu
- Office hour: Friday 1pm – 2pm

UPI:

Jorge L. Alberto: jorgealberto@ufl.edu

Andy Tran: aa.tran@ufl.edu

Kenyan M. Paschall: kenyan.paschall@ufl.edu

TA office hour: To be announced this week.

What is data science?

"Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured and unstructured data."

• Hayashi, Chikio (1 January 1998). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann; Ohsumi, Noboru; Tanaka, Yutaka; Baba, Yasumasa (eds.). Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1_3. ISBN 9784431702085

What is data science?

- Apply mathematics, statistics, and the scientific method

- Use a wide range of tools and techniques for evaluating and preparing data

- Extract insights from data using predictive analytics and artificial intelligence (AI), including machine learning and deep learning models

- Write applications that automate data processing and calculations

- Tell—and illustrate—stories that clearly convey the meaning of results to decision-makers and stakeholders at every - level of technical knowledge and understanding

- Explain how these results can be used to solve business problems

What tools will we use to perform data science?
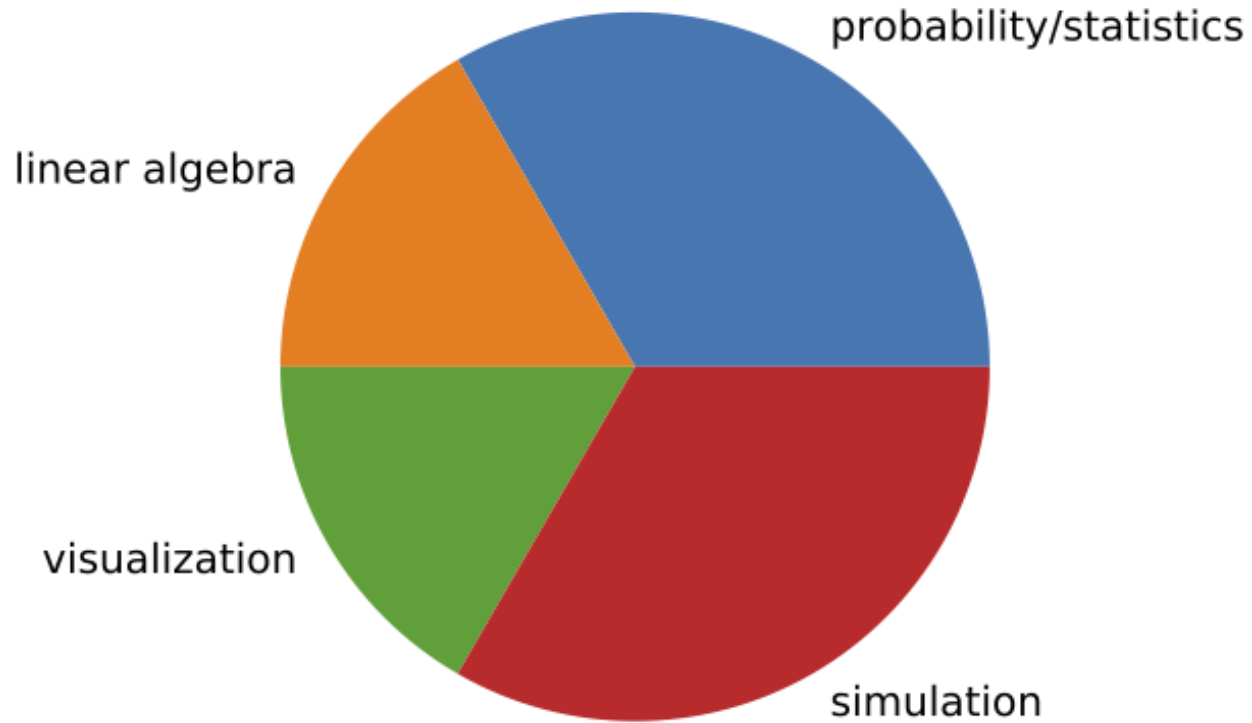
Course homepage

- [Canvas page](): announcements, assignments, grades, send/receive emails

- Slack: (invitation link:) *Expires in 27 days, signup ASAP.*
- [https://join.slack.com/t/eel3850s25/shared_invite/zt-2xftk94yc-oSuWHIHmOTGTSioK~V4oQQ]()

- Question and answering on assignment issues, general questions to TAs.

- Git Repo: [https://github.com/jiefu2017/eel3850S25]()
- Lecture notes, slides, edited lecture notes, Jupyter notebooks.

Make sure to pull the latest lecture notes prior to classes.

Course Objectives

Upon completion of this course, the student will be able to:

1. Implement, debug, and deploy Python code

2. Generate visualizations to expose meaning in data

3. Generate and understand the meaning and uses of summary statistics of data

4. Model random phenomena using random variables

5. Generate random variables with specified densities or distributions

6. Conduct hypothesis tests using simulations and analysis

7. Understand and use conditioning to simplify problems

8. Estimate parameters of distributions from samples

Course objectives

9. Understand dependence and independence among random phenomena

10. Use statistical tests to determine or characterize dependence among random phenomena

11. Design experiments to understand random phenomena

12. Understand the difference between Bayesian statistics and classical statistics

13. Use simulation to calculate Bayesian statistics

14. Apply linear algebra for data processing and statistical calculations

Time commitment

Work	Hours/Week
Attend lectures, ask questions	~4
Study/Read ~20 pages of lecture notes and code	3
Reading assignments	1.5
Homework exercises	3
Total	**11.5**

Software

- We will use Jupyter notebook in Anaconda environment: https://www.anaconda.com/download

- It includes all libraries, modules and tools we will use: Jupyter notebooks, NumPy, Matplotlib, SciPy, Pandas, scikit-learn, random

Install additional packages:

Conda: https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#creating-an-environment-with-commands

Or Pip:

https://packaging.python.org/en/latest/guides/installing-using-pip-and-virtual-environments/

References

1.Introduction to Probability

- Author: Dimitri P. Bertsekas, John N. Tsitsiklis

- Edition: 2nd

- Publisher: Athena Scientific

- Year: 2008

- ISBN: 978-1-886529-23-6

- An e-book version will be cheaper and is perfectly fine for this course. The authors created an instructional digital version of the book and you can download it : https://www.vfu.bg/en/e-Learning/Math--Bertsekas_Tsitsiklis_Introduction_to_probability.pdf

References
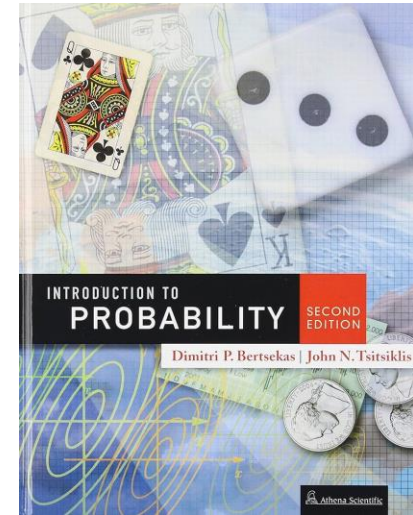
2. Introduction to Applied Linear Algebra - Vectors, Matrices, and Least Squares

- Author: Stephen Boyd

- Edition: 1st

- Publisher: Cambridge University Press

- Year: 2018

- ISBN: 978-1-886529-23-6

- An **e-book version** is freely available and is perfectly fine for this course: you can download it:

- https://web.stanford.edu/~boyd/vmls/

Stephen Boyd
Lieven Vandenberghe

Introduction to
Applied Linear Algebra
Vectors, Matrices, and Least Squares
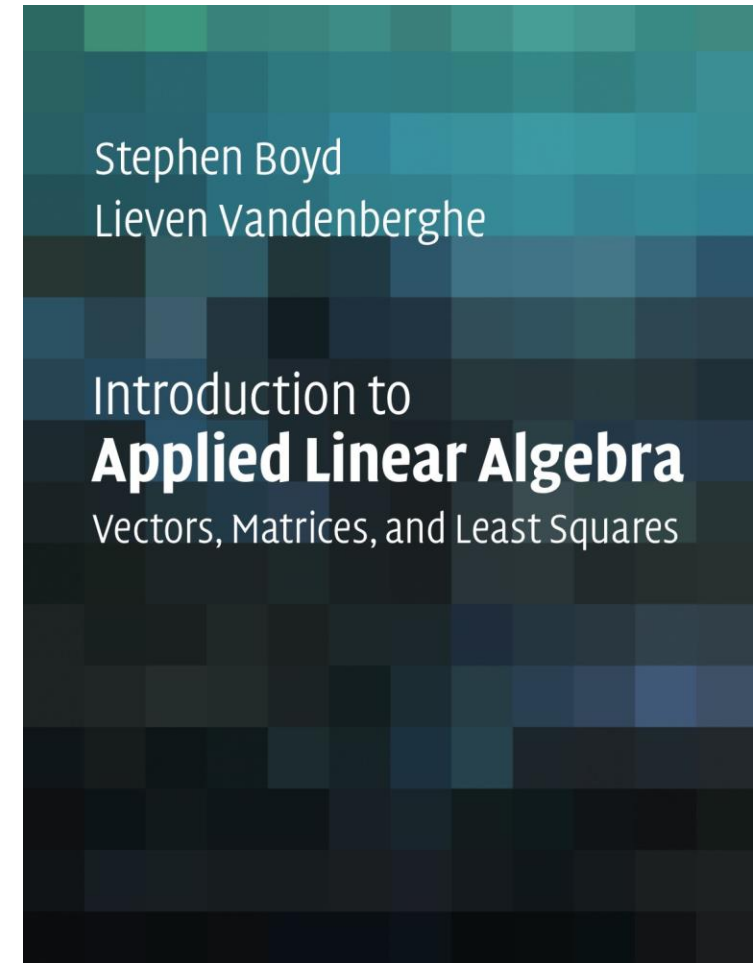
Demonstrations using Git and Jupyter Notebook

In class