INFO 6105 Data Science engineering Methods and Tools
FALL 2024

Course Information
Course Title: Neural Methods
Course Number: INFO6105
Term and Year: Fall 24
Credit Hour: 4
CRN: Click or tap here to enter text.
Course Format: Click or tap here to enter text.

Instructor Information
Full Name: Dino Konstantopoulos
Email Address: dino.k@northeastern.edu
Office Hours: TBD

Instructor Biography
Dino worked for a US Federal Lab for a decade and a half before joining Northeastern in 2018. Dino specializes in Data Science and Machine Learning, Natural Language, Algorithms, and DevOps. When Dino isn't in a class or tinkering on his laptop, he is usually on a very slim boat on the river Charles thinking about old problems and new solutions.

Teaching Assistant Information
Full Name: Dandan Zhu, Pei-Han Hsu, Yaohong Xiang
Email Address: zhu.dand@northeastern.edu, hsu.p@northeastern.edu, xiang.yao@northeastern.edu
Office Hours:TBD

Course Prerequisites
Please review the academic catalog to identify any course prerequisites

Course Description
You will learn the mathematical and algorithmic foundations of Data Science and its computational tools, while mastering Python. You will learn how to use multi-dimensional arrays and think in vectors, matrices, and tensors. You will learn how to operate on time series, spreadsheets, and higher dimensions. We will cover the basic pillars of Data Science, starting from probability theory to mathematical statistics, computational statistics with Bayesian estimation, linear algebra, and graph theory, by leveraging the 5 basic Data Science libraries written for Python: NumPy, Pandas, SciPy, Scikit-learn, and PyMC3/PyMC. This class gives you the fundamental knowledge for applying for jobs that involve data analysis, such as jobs in the life sciences, financial, advertising, and social Web industries, and to prepare you for classes in Machine Learning (ML).

Numpy adds Python support for large multi-dimensional arrays and matrices, along with a library of high-level mathematical functions to operate on these arrays. It focuses on fast number calculations, reads in fixed datatypes, improves RAM efficiency, and teaches you to think in Vectors. Pandas adds support for more refined data manipulation and analysis. It adds support for data structures and operations for manipulating tables and time series. SciPy is a collection of classic math and science algorithms and helper functions built on top of Numpy, such as linear and nonlinear regression, numerical optimization, etc. If you know the rules for dealing with your data, SciPy is the library for you. If you want the computer to learn the rules instead, and give you probabilistic answers, then Scikit-learn, built on top of SciPy, is what you need. It is a Python module for machine learning at a basic level. If you can solve a problem with the methods in SciPy, it's more straightforward. If you can't, there's a good chance your problem is solvable using methods from Scikit-learn. PyMC3/PyMC is a a popular package for running probabilistic simulations to match empirical datasets with statistical models in order to uncover their defining parameters.

Fear not if you do not have an extensive background in Math, because programming libraries will come to the rescue and do the difficult work for us. And the stuff you need to know will be up to the instructor to teach you in a manner that sticks and is entertaining.

The basic languages of Data Science are R and Python. We will start with a one-class introduction of R to get used to manipulating spreadsheets instead of single values like on a calculator, and then we use Python for the rest of the semester. A programming background in one managed language (C#, Java, or Python) is required, otherwise this class will be extremely hard for non-programmers.

Grade is based on homework (30%), midterm (30%), a final project (30%), and a final exam (10%).

Course Learning Outcomes

This class gives you the fundamental knowledge for applying for jobs that involve data analysis, such as jobs in the life sciences, financial, advertising, and social Web industries, and to be able to advance to specialized topics in Machine Learning (ML). Your knowledge of Python will improve to black-belt level.

Required Tools and Course Textbooks.

Mathematics for Machine Learning, ISBN: 110845514X, ISBN13: 9781108455145, by Marc Peter Deisenroth (Author), A. Aldo Faisal (Author), Cheng Soon Ong (Author)

Course Schedule/Topics Covered.

Introduction to Data Science
Introduction to R
Introduction to Python
Introduction to the libraries of Data Science: Numpy, Pandas, Scipy, and Scikit-Learn
Probabilities
Statistical Distributions, the p-Value, and the T-test
Frequentist Statistics and the method of Moments
Bayesian Statistics and Maximum Likelihood Estimation
Computational Statistics, the Metropolis Algorithm, and PyMC3
Time series, Regressions, and Classifications
Linear Algebra, Matrices, and Principal Components
The Calculus of Machine Learning

Assignment Grading
The grading in the class is divided up as follows:
Assignments 30%
Mid-Term Exam 30%
Final Projects (teams) 30%
Final Exam 10%

Grading Scale

	87-89.9% B+	77-79.9% C+	
95-100% A	84-86.9% B	74-76.9% C	
90-94.9% A-	80-83.9%B-	70-73.9% C-	69.9% or below F

Attendance/Late Work Policy

Attendance Policy

Students registered in MGEN courses (INFO, CSYE, and DAMG) are allowed **a maximum of 2 absences per course, with 3 or more absences resulting in an automatic 'F' for that course.** Students are expected to inform their instructors of any absences in advance of the class; if a student is sick long-term or experiences a medical issue that prevents class attendance, it is strongly encouraged that they speak with their Academic Advisor (coe-mgen-gradadvising@northeastern.edu) to learn more about the Medical Leave of Absence. Should a student anticipate being unable to attend 3 or more classes, they should discuss their situation with their Academic Advisor to explore other types of leave in accordance with the University's academic and global entry expectations. International students should review the Office of Global Services webpage to understand their visa compliance requirements.

Teaching Assistants (TAs) or Instructional Assistants (IAs) will be present at each class to collect student attendance.

Late Work Policy

Students must submit assignments by the deadline in the time zone noted in the syllabus. Students must communicate with the faculty prior to the deadline if they anticipate work will be submitted late. Work submitted late without prior communication with faculty will not be graded.

End-of-Course Evaluation Surveys
Your feedback regarding your educational experience in this class is particularly important to the College of Engineering. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this course, please take the time to complete the evaluation survey at **https://neu.evaluationkit.com**. Your survey responses are **completely anonymous and confidential**. For courses 6 weeks in length or shorter, surveys will be open one week prior to the end of the courses; for courses greater than 6 weeks in length, surveys will be open for two weeks. An email will be sent to your Northeastern University Mail account notifying you when surveys are available.

Academic Integrity
A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to http://www.northeastern.edu/osccr/academic-integrity-policy/ to access the full academic integrity policy.

MGEN Student Feedback
Students who would like to provide the MGEN unit with anonymous feedback on this particular course, Teaching Assistants, Instructional Assistants, professors, or to provide general feedback regarding their program, may do so using this survey: https://neu.co1.qualtrics.com/jfe/form/SV_cTlAbH7ZRaaw0Ki

University Health and Counseling Services
As a student enrolled in this course, you are fully responsible for assignments, work, and course materials as outlined in this syllabus and in the classroom. Over the course of the semester if you experience any health issues, please contact UHCS.

For more information, visit https://www.northeastern.edu/uhcs.

Student Accommodations
Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit https://drc.sites.northeastern.edu.

Library Services
The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for education specific resources, visit https://library.northeastern.edu
Network Campus Library Services: Northeastern University Library Global Campus Portals

24/7 Canvas Technical Help
For immediate technical support for Canvas, call 617-373-4357 or email help@northeastern.edu

Canvas Student Resources: https://canvas.northeastern.edu/student-resources/

For assistance with my Northeastern e-mail, and basic technical support:
Visit ITS at https://its.northeastern.edu
Email: help@northeastern.edu
ITS Customer Service Desk: 617-373-4357

Diversity and Inclusion
Northeastern University is committed to equal opportunity, affirmative action, diversity, and social justice while building a climate of inclusion on and beyond campus. In the classroom, members of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration, and an awareness of global perspectives on social justice.

Please visit http://www.northeastern.edu/oidi/ for complete information on Diversity and Inclusion

Title IX
Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty, and staff.

In case of an emergency, please call 911.

*Please visit **https://www.northeastern.edu/ouec** for a complete list of reporting options and resources both on- and off-campus*.