

Fall 2023

DS 675-005: Machine Learning

Mengjia Xu

Follow this and additional works at: <https://digitalcommons.njit.edu/ds-syllabi>

Recommended Citation

Xu, Mengjia, "DS 675-005: Machine Learning" (2023). *Data Science Syllabi*. 4.
<https://digitalcommons.njit.edu/ds-syllabi/4>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Data Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

New Jersey Institute of Technology
Ying Wu College of Computing
Data Science Department

Machine Learning

Code: DS675

Time: Friday, 2:30 PM–5:20 PM, Fall 2023

Location: ECEC 115

Mode: Face-to-Face

Instructor: Mengjia Xu

Office: GITC 3901E

Email: mx6@njit.edu

Office Hours: Mon (1:45pm-3:10pm), Tue (1:00pm-1:40pm), Thu (12:15pm-12:55pm, [Webex](#)). If these hours do not work with your schedule, appointments are also available by email.

Teaching Assistant/Grader: TBD

Note: Your messages will be answered by the end of the next day. Grades for all items will be getting posted during the week after their due date. For issues with your grades, contact the grader and cc the instructor.

Tutoring. NJIT provides a [tutoring service](#). Please contact one of the available tutors. Please check the website for updates regularly as they may change the information.

Course Description

[From NJIT catalog]: This course is an introduction to machine learning and contains both theory and applications. Students will get exposure to a broad range of machine learning methods and hands-on practice on real data. Topics include Bayesian classification, perceptron, neural networks, logistic regression, support vector machines, decision trees, random forests, boosting, dimensionality reduction, unsupervised learning, regression, and learning new feature spaces. There will be several programming assignments, one course project, one mid-term and one final exam.

[Instructor's description]: Machine Learning develops computer programs that can improve their performance by tapping into existing data and taking feedback from the environment. Systems based on ML have already exceeded human performance in several tasks, including image medical image classification and games like Chess and Go. ML has also made leaps in even more complicated tasks, like Natural Language Processing or self-driving vehicles, and it has even produced art that imitates the style of human artists! This course offers an intense introduction to the fundamental ML concepts and algorithms that constitute the core of these spectacular developments. It takes you on a tour from the basic mathematical notions and algorithms to some of the recent developments, e.g., deep convolutional neural networks, recurrent neural networks, transformers. You will gain exposure to cutting-edge ML development tools such as Jupyter notebooks, scikit-learn, and PyTorch via hands-on

assignments and projects that will instill a working and immediately applicable knowledge of ML methods and will prepare you for more advanced ML courses.

Prerequisites

The course does not have other course prerequisites.

Background on some basic calculus, linear algebra, probability and programming ability is required. The following free online materials are recommended for reviewing this background:

- [Mathematics for Machine Learning](#)
 - [A visual guide to NumPy](#)
-

Course Textbooks

There is no required course textbook. The course will draw material from several sources, including the instructor's own notes. Some optional resources include:

- Bishop, Christopher M., 2006. [Pattern Recognition and Machine Learning](#) (PRML). Springer-Verlag New York, Inc. A comprehensive reference for Bayesian theory that we will cover.
 - Ian Goodfellow and Yoshua Bengio and Aaron Courville, 2016. [Deep Learning](#) (DL), MIT Press. We will cover topics including basic neural networks, back propagation, and CNN.
 - Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome, 2008. [The Elements of Statistical Learning](#) (ESL). Second Edition, Springer New York Inc.
 - Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola, 2021. [Dive into Deep Learning \(DDL\)](#). arXiv preprint arXiv:2106.11342.
 - Haul Daume III. [A course in Machine Learning](#) (Online book)
 - Aurelien Geron. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems (2nd Edition). [[GitHub](#)]
 - Sebastian Raschka and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2 (3rd edition).
-

Learning Outcomes

By the end of the course, you will be able to:

- Be more proficient at Python programming
- Identify the main types of Machine Learning (ML)
- Describe and explain a wide variety of ML algorithms
- Apply various ML algorithms in novel situations
- Evaluate the performance of ML models
- Modify ML models to improve their performance
- Adapt ML algorithms and models to the given data and application

Coursework, Assessment and Related Outcomes

Homework [30%]. Six homework assignments of equal grading weight. The weakest of the six grades is dropped from the calculation.

Interview-style Quizzes [10%]. Weekly Canvas quizzes reinforcing the material of each module.

Class Participation [10%]. You are expected to attend classes and participate in classes by listening and understanding class contents and asking related questions. You are also expected to participate in weekly Canvas discussions prompted by the instructor, with meaningful questions and answers related to the week's topics or assignments.

Midterm Exam [15%]. In-person exam, 80 minutes. Students are expected to bring a fully charged laptop, as the exam will be on Canvas with LockDown browser. Each student is allowed to bring at most 5 pages of notes. In the event the exam has to take place online, Respondus Monitor will be used for proctoring.

Final Exam [15%]. Cumulative, 120 minutes. Otherwise, similar to midterm exams.

Mini-Project [20%]. You are expected to complete a mini-project in a group. The project will consist of four milestones with weights: [4%, 1%, 5%, 10%]. More details please check the project requirement document [link](#).

Grading Scheme and Letter Grades. The conversion of raw total scores to letter grades will be based on grouping the scores into clusters and then assigning a letter grade to each cluster. Projected letter grades will be communicated after the midterm exam and will be constantly updated throughout the rest of the semester, to reflect the underlying clusters. The final letter grade assignment will always be in accordance with the graduate [grade legend](#).

Grading Feedback. Assignment marks will be accompanied with solutions and general feedback summarizing common mistakes. Individual grading feedback will be given whenever possible. Further clarifications can be provided via direct communication with the instructor and the course grader.

Late Work Policy. In the case when a student is unable to complete an assignment or other serious reasons, these must be communicated and documented promptly. In any other case, each hour of delay after the due date will incur a 2% score reduction. No extensions will be granted. However, the lowest programming score and the two lowest quiz assignments scores for each student will be dropped.

Course Topic Schedule (tentative)

The topics covered in this course include the following, presented in the approximate order in which they will be taught. This list of topics is to be considered as a *reference* that can be adjusted through the course of the semester to address changing needs.

Week 1	1. Introduction and overview
---------------	------------------------------

Week 2	2. Basics of Python, Linear regression and Gradient Descent
Week 3	3. Linear Classification and Perceptron
Week 4	4. Generative Models and Logistic regression
Week 5	5. SVM and Decision Tree
Week 6	6. Ensembles and Boosting
Week 7	7. Dimensionality reduction, Kernel PCA; Unsupervised Clustering
Week 8	8. <i>Midterm exam</i>
Week 9	9. Artificial Neural Networks
Week 10	10. Autoencoders
Week 11	11. Convolutional Neural Networks
Week 12	12. Recurrent Neural Networks
Week 13	13. Transformers & LLMs
Week 14	Project presentations & Review

Assignment Due Dates

Each course module is associated with a 10-minute theory review quiz which is due on the end of the week when the topic is presented (*).

All other due assessment items are due at the end of the week indicated in the following schedule:

Homework [30%]	Project Milestones [20%]	Exams [30%]
#1: Week-2 #2: Week-4 #3: Week-6 #4: Week-9 #5: Week-11 #6: Week-13	#1: Week-3 #2: Week-5 #3: Week-9 #4: Week-13	Mid-semester: Week-8 End-semester: TBD

(*) All items are due on Sunday at 23:59.

Course Policies

General

- Please feel free to join office hours (with me and TA's) to discuss any issues.

- Email is the best way to get in touch with the instructor. Please include “DS 675005” in the subject line of your email.
- Please do not hesitate to contact me if you have any problems, concerns, questions, or issues regarding the course, material, or anything else in the class.
- Please do not hesitate to talk to me if there are situations in your life that are affecting your performance in the class or your life here at NJIT. I might not be able to help, but I might know of resources that might help.

Email

Use of your NJIT email or Canvas inbox is strongly encouraged.

Grade Corrections

Check the grades in course work and report errors promptly. Please try and resolve any issue within one week of the grade notification.

Exam and Proctoring Policy

See the [NJIT Online Course Exam Proctoring page](#) for information on proctoring tools and requirements.

Incomplete

A grade of I (incomplete) is given in rare cases where work cannot be completed during the semester due to documented long-term illness or unexpected absence for other serious reasons. A student needs to be in good standing (i.e., passing the course before the absence) and receives a provisional **I** if there is no time to make up for the documented lost time; an email with a timeline of what is needed to be done will be sent to the student. Note that an **I** must always be resolved by the end of the next semester.

Fail of the Course

In the case when a student is unable to attend the class or exams, these must be communicated and documented promptly. Unless emergency, a student should communicate with the instructor prior to the class or exam time. In any other case, a student will fail this course and obtain an F if 1) missing more than three classes; 2) missing any exams; 3) not submitting course project final report. No exceptions will be granted.

Collaboration and External Resources for Assignments

Some homework problems will be challenging. You are advised to first try and solve all the problems **on your own**. For problems that persist you are welcome to talk to the course assistant or the instructor. You are also allowed to collaborate with your classmates and search for solutions online. But you should use such solutions only if you understand them completely (admitting that you don't understand something is way better than copying things you don't understand). Also make sure to give the appropriate credit and citation.

Generative AI Tools and Other External Resources

Sometimes you may come across code, text or other helpful information online, or you may be able to generate it using AI tools such as ChatGPT or other Large Language Models (LLMs). In most cases, you will be allowed to integrate this information into your solution. However, if you do, you must always give the appropriate credit and citations (e.g. links) for the material you use (especially when you use the code and text you find online). In the case you use an LLM, you must say that you did so, and present the entire transcript of your 'conversation' with it, which should show what you asked and how you guided it, or were guided by it to the delivered solution. Your 'conversation' with it must be entirely yours, and sufficiently different from that of other students. Failure to give appropriate credit when using the work of others (whether human or AI) is considered plagiarism, and may lead to disciplinary action under NJIT's Academic Integrity policy (see below).

Requesting Accommodations

If you need accommodation due to a disability please contact Scott Janz, Associate Director of the [Office of Accessibility Resources and Services](#), Kupfrian Hall 201 to discuss your specific needs. A Letter of Accommodation Eligibility from the office authorizing student accommodations is required.

NJIT Services for Students, Including Technical Support

Please follow this [link](#).

Canvas Accessibility Statement

Please follow this [link](#).

Academic Integrity

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at this [link](#).

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. **Any student found in violation of the code by cheating, plagiarizing, or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university.** If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu.