Fall 2023

CS 375: Introduction to Machine Learning

Lijing Wang

Updated: 09/05/2023

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

Machine Learning

Code: CS375
Time: Tuesday/Thursday, 11:30AM–12:50PM, 2023 Fall
Location: Cullimore Hall LECT 2
Mode: Face-to-Face

Instructor: Lijing Wang
Office: GITC 5715
Email:lijing.wang@njit.edu
Office Hours: Tuesday/Thursday 1:45 pm to 2:25 pm, 2:30 pm to 3:10 in person. Wednesday 1:45 pm to 2:25 pm online by appointment. Reserve an online appointment slot by following this calendar link. Please try to do so at least one day in advance. If these hours do not work with your schedule, appointments are also available by email.

Teaching Assistant with Office Hours:

- Koushik Chandrasekaran, kc664@njit.edu, Thursday 3:00 pm to 4:00 pm (may change) via Webex: https://njit.webex.com/meet/kc664

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 Networks or Recurrent Networks. You will gain exposure to cutting-edge ML development tools such as 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:

a. Identify the main types of Machine Learning (ML)
b. Evaluate the quality of online resources related to ML
c. Recognize problems amenable to ML methods
d. Describe and explain a wide variety of ML algorithms
e. Apply various ML algorithms in novel situations
f. Evaluate the performance of ML models
g. Modify ML models to improve their performance
h. Adapt ML algorithms and models to the given data and application

---

Coursework, Assessment and Related Outcomes

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

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 and is subject to change)

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 a *reference* that can be adjusted through the course of the semester to address changing needs.

Week 1	Introduction and Overview	Ch1-2 PRML 2006 [Salimans et al 2016] [Reed et al 2016] [Graves et al 2013] [Google Brain 2016] [Farfade et al 2015]
Week 2	Linear Regression	Ch3 PRML 2006 Ch3.1 DDL 2021
Week 3	Linear Classification	Ch4 PRML 2006 Ch3 ESLII 2008 Ch4.1 DDL 2021
Week 4	Logistic Regression PCA	Ch4 PRML 2006 [McMahan et al 2013]
Week 5	Support Vector Machines	Ch7 PRML 2006 [Cortes& Vapnik 1995]
Week 6	Decision Trees Ensembles and Boosting	[Quinlan 1985] [Friedman 2001] [Chen et al 2016]
Week 7	Nonparametric Methods Unsupervised Clustering	Ch2.5 PRML 2006
Week 8	Midterm Review, Q&A Midterm Exam	

Week 9	Graphical Models	Ch8 PRML 2006
Week 10	Introduction to Neural Networks	Ch5 PRML 2006
Week 11	Tips for Deep Learning	
Week 12	Convolutional Neural Networks	Ch9 DL 2016 [LeCun et al 1998] Ch7 DDL 2021
Week 13	Recurrent Neural Networks	Ch10 DL 2016 Ch9 DDL 2021
Week 14	Project Presentation	
Week 15	Final Review, Q&A	
Week 16	Final Exam	

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 [25%]	Interview-style Exams [25%]
#1: Week-2 #2: Week-4 #3: Week-6 #4: Week-9	#1: Week-3 #2: Week-5 #3: Week-9 #4: Week-13	Mid-semester: Week-8 End-semester: TBD

#5: Week-11 #6: Week-13		

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

Course Policies

General

- Please feel free to join the 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 675002" 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.

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.