Fall 2024

DS 675-851, 853, 855: Machine Learning

Michael Houle

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

Machine Learning

Code: FA24-DS675-851/853/855
Mode: Online

Instructor: Michael Houle
Webpage: https://people.njit.edu/faculty/meh43
Office: GITC 4317D (Newark)
Email: michael.houle@njit.edu (or directly on Canvas)

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

Office Hours: Tue 5:00pm–7:00pm & Thu 6:00pm–8:00pm, online (Zoom).
Reserve an online appointment slot by following this calendar link. Please try to do so at least one day in advance. Appointments can also be arranged by email.

Extra Hours: Optional Class Meetings on Thu 4:00pm–5:00pm OR Thu 9:00pm–10:00pm, online (Zoom). The Class Meetings are live sessions in which students can ask questions on the material covered in the modules. Occasionally, extra material may be presented that expands upon the weekly topics. Students can freely choose which of the two Meetings to attend, but should attend only one per week. Attendance at a Class Meeting is not mandatory, but is strongly recommended. Videos of all Class Meetings will be made available in Canvas.

Teaching Assistant / Grader: Hamideh Sabaei (hs833@njit.edu).

Tutoring: NJIT provides a tutoring service. Tutor availability and contact information can be found here.

Course Description

[From the 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 (ML) 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 complex tasks, such as Natural Language Processing or self-driving vehicles, and generating text, code images, video and music almost indistinguishable from that produced by human experts and artists. This course offers an intensive 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 the foundations of recent developments in Deep Learning (DL). 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 which will prepare you for more advanced courses in ML and DL.

Prerequisites

This course does not have other courses as prerequisites.

Background on basic calculus, linear algebra, probability and statistics, and Python programming is required. Familiarity with the following mathematical concepts and their notation is essential for success in this course.

Calculus	Linear Algebra	Probability and Statistics
Basic notion of derivatives	Vectors and vector notation	Basic probability and selection
Differentiation of polynomials, exponentials, logarithms	Lines, planes, hyperplanes	Conditional probability & independence
	Normal vectors	Bayes' theorem*
	Vector norm	Maximum likelihood*
	Dot product	Random variables
Product rule	Orthogonality and projection	Expectation
Quotient rule	Linear transformation	Mean, variance, deviation
Chain rule	Matrix notation	Error, bias
Partial differentiation	Matrix addition & multiplication	Standardization
Multivariate chain rule	Matrix transpose, inverse	Distributions and sampling
Gradient	Matrix rank	Probability density function
Basic notion of integrals	Eigenvalues and eigenvectors*	Cumulative distribution function
Fundamental Theorem of Calculus	Basis vectors*	Uniform distribution
	Eigendecomposition*	Normal distribution (Gaussian)

* = A review of this topic will be presented for those who have not encountered it before.

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 on material from several sources, including the instructor's own notes. Some optional resources include:

- A Course in Machine Learning
  Hal Daumé III — Online book (draft)

- Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow
  Concepts, Tools and Techniques to Build Intelligent Systems (3rd edition)
  Aurélien Géron — O'Reilly, ISBN-13: 978-1098125974

- Python Machine Learning: Machine Learning and Deep Learning with Python,
  scikit-learn, and TensorFlow 2 (3rd edition)
  S. Raschka, V. Mirjalili — Packt Publishing, ISBN-13: 978-1789955750

- The Elements of Statistical Learning (2nd edition)
  T. Hastie, R. Tibshirani, J. Friedman — Springer, ISBN-13: 978-0387848570

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 in order to improve their performance.
h. Adapt ML algorithms and models to the given data and application.

Coursework, Assessment, and Related Outcomes

Participation [10%]
You are expected to participate in class activities throughout the term. The main venues for this are the weekly Canvas discussion forums. Each week you should make at least two substantial contributions that are of clear benefit to your peers.
[Outcomes: a-d]

Short Theory Reviews [10%]
20 weekly Canvas quizzes reinforcing the material of each module, with no browser restrictions, and **open** books/notes/web.
[Outcomes: a,d]

Programming Assignments [25%]
Six hands-on Machine Learning programming assignments.
[Outcomes: c,e-h]

Mini-Project [25%]
An implementation project in Machine Learning consisting of four milestones with the following weights: [3%, 6%, 1%, 15%].
[Outcomes: b,c,e]

Professional Quiz (Mid-Semester) [10%]
A 90-minute online quiz of the style often seen by ML professionals in a job interview, covering all material in Weeks 1-7. Each student is allowed to bring at most 5 pages of personally-handwritten notes. These must be uploaded to Canvas before the exam. The exam will be on Canvas with proctoring using Respondus LockDown browser with video recording. Students must ensure that their computer is equipped with a working webcam. Students must come to the exam with a fully-charged laptop.
[Outcomes: a,d,e]

Professional Quiz (End-of-Semester) [20%]
Covers the entire course. 120 minutes. In other respects, similar to the Mid-Semester Quiz.
[Outcomes: a,d,e]

Course Topic Schedule

Week 1	1. Introduction to DS675 Machine Learning 2. k-Nearest Neighbors and Basic ML Notions
Week 2	3. Linear Regression, Gradient Descent and Feature Engineering
Week 3	4. Perceptron and Linear Separability
Week 4	5. Logistic Regression 6. Regularization
Week 5	7. Support Vector Classifiers 8. Decision Trees
Week 6	9. Hyperparameter Tuning and Validation 10. Attribute Selection
Week 7	11. Ensembles and Boosting
Week 8	12. A Probabilistic Perspective for ML Algorithms (midterm week)
Week 9	13. Dimensionality Reduction and Kernel PCA 14. Unsupervised Clustering
Week 10	15. Introduction to Artificial Neural Networks 16. Nonlinear Layers, Softmax and Cross-Entropy
Week 11	17. Autoencoders 18. Introduction to PyTorch
Week 12	19. Convolutional Neural Networks
Week 13	20. Recurrent Neural Networks
Week 14	Project Presentations

Assignment Due Dates

Each course module is associated with a 10-minute theory review quiz, due at the end of the week when the topic is presented. All other assessment items are due in the week indicated in the following schedule.

Assignments [25%]	Project Milestones [25%]	Professional Quizzes [30%]
#1: Week 2 [2.5%] #2: Week 4 [4.5%] #3: Week 6 [4.5%] #4: Week 9 [4.5%] #5: Week 11 [4.5%] #6: Week 14 [4.5%]	#1: Week 3 [3%] #2: Week 7 [6%] #3: Week 9 [1%] #4: Week 14 [15%]	Mid-Semester: Week 8, Sat. evening [10%] End-of-Semester: Final exam period [20%]

The deadline for all assessment items other than Project Milestone 4 and the Professional Quizzes is at 23:59 on Sunday at the end of the week they are due. Project Milestone 4 is due mid-week, at 23:59 on Wednesday.

Grading Policies

Letter Grades
In accordance with the graduate grade legend, the raw total percentage assessment score will be converted to a final letter grade that will appear on your transcript. The conversion table for this course is:

Letter Grade	Percentage Range
A	90 — 100
B+	80 — <90
B	70 — <80
C+	65 — <70
C	60 — <65
F	<60

In exceptional cases the letter grade may be raised one level higher than what is suggested by the above conversion table. Examples include a top 5%-in-class score in the proctored exams, or a project with very significant novel contributions.

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 requesting special consideration should be in good standing (i.e. with a passing grade on coursework submitted before the absence). When special consideration is granted, the student receives a provisional **I** if there is no other way to make up for the documented lost time; in such

cases, an email with a timeline for makeup work will be sent to the student. Note that according to NJIT regulations, an **I** must always be resolved by the end of the following semester.

Late Submission Policy
Generally speaking, assignments, project milestones, and participation exercises will be accepted late without penalty, but only up until the time grading has begun, or solutions or other feedback have been released to members of the class. At that time, the class will be informed that submissions have closed, that no further submissions will be accepted, and that any missing work will be given a mark of zero. Students should be aware that grades, solutions, and/or feedback may be released at any time after the deadline, at the sole discretion of the lecturer and without prior warning. *The only way to ensure that work will be accepted without penalty is to submit it before the deadline*. No individual extensions will be granted. However, special consideration may be given in rare cases when a student is unable to complete an assignment for serious, unavoidable reasons — these must be communicated and documented promptly.

Grading Feedback
Assignment marks will be accompanied with solutions and/or general feedback summarizing common mistakes. Individual grading feedback will be given where appropriate. Further clarifications can be provided by contacting the instructor and/or the course grader.

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.

Other Course Policies

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

Exam and Proctoring Policy
See the NJIT Online Course Exam Proctoring page for information on proctoring tools and requirements.

Requesting Accommodations
If you need an accommodation due to a disability, please contact Marsha Williams-Nicholas, 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.

Collaboration on Assignments
Some of the assignment problems will be quite challenging. You are advised to first try to solve all the problems **on your own**. For problems that persist you are welcome to talk to the instructor during office hours, or raise questions in the Weekly Discussion Forum. In consulting with others, you are allowed to exchange general ideas and approaches only: unless you are given explicit permission to do so in the assignment statement, the full solutions themselves must be worked out by you alone.

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 found 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).

Statement on 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: http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf .

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 ."*