

DSE 6211: Machine Learning + Artificial Intelligence

Instructor: Peter Salemi, Ph.D.

Instructor Email: salemip@merrimack.edu

Course Term: Spring Session 1 (Jan-15-2024 to Mar-8-2024)

Synchronous Live Session: Tuesdays, 8-9PM ET

Office Hours: Monday - Friday, by appointment

Course Description

This course offers students an introduction to machine learning and artificial intelligence, focusing on supervised and unsupervised learning. More specifically, students will learn the proper machine learning workflows and techniques for extracting critical insights from data, and then apply their knowledge to real-world data sets. The course is meant to offer an overview of this highly complex field, focusing on key intuition behind established approaches and hands-on applications of techniques. Topics include data pre-processing, regression and classification using deep learning models, evaluation of classifiers, principal component analysis, and clustering.

Credit Justification Statement

In this course, students will be expected to do the following amount of work:

- Live session: 1 hour per week
- Readings: 8 hours per week (on average)
- Labs: 8 hours per week (on average)
- Course assignments: 6 hours per week (on average)
- Total: 23 hours per week (on average)

Course Learning Objectives

Upon completion of this course, students should:

- Understand the general processes and methodologies comprising the field of *machine learning*, including *data pre-processing*;
- Understand data, process, and outcome related differences distinguishing *unsupervised* and *supervised* modes of machine learning;
- Develop basic understanding of the key analytic differences distinguishing *classification* and *regression*;
- Understand how to select the proper architectures and components of *deep learning models*, as well as how to train and evaluate deep learning models in practice;
- Understand when and how to implement *principal component analysis* to potentially reduce the dimension of a dataset.
- Demonstrate rudimentary competency in conducting and evaluating basic machine learning analyses.

Important Deadlines

Jan-15-2024	First day of classes
Jan-19-2024	Last day to add/drop a course
Feb-26-2024	Last day to withdraw from a course with a “W”
Mar-8-2024	Last day of classes

Required Course Materials (see the links on the Canvas syllabus page for where to get them):

1. (DLwR) Chollet, F., Kalinowski, T., and Allaire, J. J. *Deep Learning with R*. (Second Edition). Manning.
2. (ISL) James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. (Second Edition). Springer.

Software

The instructor will utilize R to illustrate all techniques during the instructional process. However, students fluent in Python can use it to complete the assignments.

Grading & Assignments

A Excellent	B Good	C Fair/Poor	F Unacceptable
93+ A	87-89 B+	77-79 C+	Below 70
90-92 A-	83-86 B	73-76 C	
	80-82 B-	70-72 C-	

Although the graduate grading policy is similar to the undergraduate policy, it should be noted that the expectations for graduate students are much higher and therefore the grading is more rigorous. Candidates for any graduate-level degree or certificate must attain a final cumulative grade point average of 3.0 before the degree or certificate will be conferred.

The following letter and special grades are used across all graduate programs. Please note that this document sets the minimum standards for the College. Individual degree programs may have stricter GPA and course grade requirements.

A: “A” indicates outstanding work

B: “B” means that the work is satisfactory

C: “C” (2.0) is deemed unsatisfactory at the graduate level. No more than two courses at the C level (2.0 or higher) will be counted as acceptable toward a graduate degree. Students may be permitted to repeat only two courses and may repeat each course only one time. Those who receive more than two C level grades will be automatically dismissed from their program of study.

Any grade lower than a C - will not be acceptable for graduate-level work and cannot be counted as credit towards the degree. However, the grade will be counted toward the graduate student’s GPA. If a student receives a grade lower than a C in a required course, the student will be required to retake the course. A course may only be retaken one time. The most recent grade will replace the first grade and will be factored into the student’s GPA.

Please see “Academic Requirements and Policies” in the Graduate Catalog at <https://catalog.merrimack.edu/index.php> for more information.

Student requirements and their corresponding weights are summarized as follows:

Assignment	Weight
Analytic Plan	10%
Preliminary Results	20%
Final Report	30%
Labs (8)	40%

(1) Analytic Plan, Preliminary Results, and Final Report

The main purpose of the course project is to work through the full lifecycle of a relatively simple, yet very realistic, applied machine learning project. With this in mind, the project is divided into three general components: 1) Analytic Plan, 2) Preliminary Results, and 3) Final Report.

The Analytic Plan is a review for internal or external clients containing the data scientists' prepared plan of analysis, presented in the context of a stated business need and data to be used.

The Preliminary Results is a review of the preliminary (i.e., initial or intermediate) machine learning models, the goal of which is to evaluate the models and determine possible areas for improvement. With this in mind, the assignment is focused primarily on hands-on analysis, which encompasses the data preparation steps and supervised learning tasks outlined in the Analytic Plan.

The Final Report is the final client deliverable, where a data scientist presents the earlier reviewed, and amended, if necessary, results and recommendations, all presented in the context of the key elements of the Analytic Plan, emphasizing the stated goal(s) of the project. As such, the Final Report assignment is characterized by greater scope (than the Analytic Plan and Preliminary Results assignments) as it needs to address the key elements of both the Analytic Plan and Preliminary Results, in addition to also delineating the key takeaways and recommendations.

(2) Labs

There are 8 Labs throughout the course, each providing hands-on experience implementing and reinforcing the techniques and methods discussed during the Live Session and Readings. Students are expected to prepare and submit responses to the exercises found at the end of each Lab to receive credit. **FOR UNDERGRADUATE STUDENTS ONLY: only 6 of the 8 labs are required.**

Grading-Related Policies for this Course

All assignments must be submitted through Canvas by 11:59PM of the due date listed in the Weekly Topics & Assignments section and in Canvas. Please note that there will be **NO late work accepted**. Final grades will be based on a percentage of possible points earned, but the instructor reserves the right to adjust the course grade.

1. Academic Integrity

See Merrimack College Student Handbook for college's Academic Integrity Policy. The academic integrity code and policy is also posted in the Graduate Catalog:

<https://catalog.merrimack.edu/content.php?catoid=9&navoid=202#academic-integrity>

Here is a brief excerpt from that: "...Academic integrity is fundamental to creating and maintaining an atmosphere of cooperation and trust. It is thus a concern for everyone in the college community. The academic integrity code below is designed to help students understand what is not permissible in their academic and intellectual lives at the college. It seeks to protect students from unintentional acts of dishonesty and to preserve the trust inherent in the student- teacher relationship, which is compromised if suspicion arises regarding the integrity of a student's work. The code is also designed to inform students of the rules which will be used to judge academic integrity infractions..."

Specifically concerning the production of code it is important to recall that using code, text or other forms of media, such as homework solutions, from previous offerings of the course is forbidden, except as a "reference" or "guide" to work from, and even as such, **it is necessary to explicitly cite all consulted sources**. Tutors or other students who have taken the course in the past may help current students, but should not show them the solutions that they have developed for assignments. The lack of detailed citation of previous offerings of the course or any other similar sources, notably from the Internet or Generative AI systems, will be considered as submission of copied material, and a serious breach of academic integrity. Simply put, any work that is not your own should be clearly identified either through citations in written work or in comments in submitted files. This includes not only things that are directly copied from other's work, but also for ideas and procedures gathered from outside sources. It is not necessary to cite the materials presented in class.

2. Academic Accommodations from the Accessibility Services Office

Merrimack College provides reasonable accommodations for students with documented disabilities. Students who have, or think they may have, a disability are invited to contact the Accessibility Services Office via the online request form found on the Accessibility Services website: www.merrimack.edu/aso, email: accessibilityservices@merrimack.edu, or by visiting us on the third floor of McQuade Library (in the event we are open).

Students are encouraged to contact the office as soon as possible to ensure adequate time to meet and create a plan. Students already registered with Accessibility Services are encouraged semesterly to request for their letters to be emailed and students are responsible to then email the PDF to their instructors personally. Accommodations cannot be made retroactively.

3. Live On-line Meeting Attendance

Participation in weekly live on-line meetings is highly encouraged, but it is not mandatory. All sessions will be recorded, and recordings will be posted; however, when not attending students forego the ability to directly ask questions about topics being discussed, or other parts of the course.

4. Communication

Class communication will come through official college email (merrimack.edu) and Canvas announcements as needed. Those will be the primary methods for dissemination of information between instructors and students, so please check your email at least once a day. Instructors check their email several times a day, so it is the quickest way to reach them with questions. Important announcements will be archived through Canvas. When emailing your instructor, use your Merrimack account, include a greeting with the class code and a closing signature with your name, or they may not respond.

If you are having trouble with an assignment, please do not suffer in silence. Let your instructor know if you are struggling, either by email or by scheduling an appointment. We are all interested in seeing students succeed, and are more than happy to provide extra help.

5. Requests for Extensions

The general policy is that, outside of properly verified serious medical emergencies* (as defined below), extensions are not given, which applies to all assignments. Missing an assignment without an acceptable reason (to be clear, that means a serious medical emergency, as defined below) will result in 0 points for the exam or a project. The intent here is not to penalize anyone – quite to the contrary, it is to create a level playing field so that no one has a unique and an unfair advantage. All assignment due dates are published (see the Live Meetings, Topics & Assignments section below) and will not change, barring a natural or other emergency – please consider those dates when planning any non-class related activities.

**Serious medical emergency is defined as an injury or illness that is acute and poses an immediate risk to a person's life or long term health. To be "properly verified", the said serious medical condition must be attested to by hospitalization and related medical treatment documentation.*

6. Mental Health

It is important to make your mental health a priority! We will do activities in class to support your mental health and wellness, and I also encourage you to engage in your own self-care habits outside of class. If you want more information or resources, please come see me! If you are struggling with your mental health, or you believe a classmate is, please talk to me so I can put you in touch with qualified and caring support to get you back on track and feeling better.

7. Suicide Prevention Lifeline

We can all help prevent suicide. The Lifeline provides 24/7, free and confidential support for people in distress, prevention and crisis resources for you or your loved ones, and best practices for professionals.

http://www.ulifeline.org/stay_well

National Suicide Prevention Lifeline 1-800-273-8255(TALK) or Call or Text 988

8. Counseling Services

Counseling Services are available to Graduate students at the Counseling Center at Merrimack via Uwill for telecounseling services only. Telecounseling includes video, chat, phone, message. Uwill is HIPPA and FERPA compliant which means they follow the strictest privacy guidelines. To access Uwill, register with your Merrimack email address at app.uwill.com. To receive support after hours call 978-837-5444 to be connected to the Uwill 24/7/365 crisis line which offers phone assistance only. The Counseling Center is located on the third floor of the Sakowich Center.

9. Student Success Resources

It takes a village to progress in your knowledge. The numerous resources available at Merrimack College to assist you in meeting your goals are outlined in the **Data Science Hub in Canvas**.

For assistance in navigating the student success resources please reach out to the graduate advising team via ecs-grad-advising@merrimack.edu

Weekly Topics & Assignments

Date	Topics	Readings & Discussions
Week 1: Introduction to Machine and Deep Learning	Traditional machine learning (ML) -Linear regression -Logistic regression Deep learning as a generalization of traditional ML -Linear regression as a neural network (NN) -Logistic regression as a NN -Automated feature engineering	DLwR chapter 1 ISL 3.1–3.2, 4.1–4.3, 10.1 Lab #1 (due: Jan. 21, 11:59PM ET)
Week 2: Fundamentals of NNs	NN architectures Tensors Training NNs (backpropagation and gradient descent)	DLwR chapter 2 ISLR 10.1–10.2 Lab #2 (due: Jan. 28, 11:59PM ET)
Assignment 1 – Analytic Plan – due: Jan. 28, 11:59PM ET		
Week 3: Tools for Building NNs	Keras Tensorflow	DLwR chapter 3 Lab #3 (due: Feb. 4, 11:59PM ET)
Week 4: Classification and Regression with NNs	Binary classification with NNs Multiclass classification with NNs Regression with NNs	DLwR chapter 4 Lab #4 (due: Feb. 11, 11:59PM ET)
Week 5: Fundamentals of ML	Underfitting and overfitting Evaluating machine learning models Regularization	DLwR chapter 5 Lab #5 (due: Feb. 18, 11:59PM ET)
Assignment 2 – Preliminary Results – due: Feb. 18, 11:59PM ET		
Week 6: ML Workflow	Developing models Deploying models Review of NNs	DLwR chapter 6 Lab #6 (due: Feb. 25, 11:59PM ET)
Week 7: Unsupervised ML Part I	Principal component analysis Scree Plots and Parallel Analysis	ISL 12.1–12.2 Lab #7 (due: Mar. 3, 11:59PM ET)
Week 8: Unsupervised ML Part II	K-means clustering Hierarchical clustering	ISL 12.4 Lab #8 (due: Mar. 8, 11:59PM ET)
Assignment 3 – Final Report – due: Mar. 8, 11:59PM ET		