

CHM4930
Machine Learning in Chemistry
Fall 2021

ALL CLASSES WILL BE HELD IN PERSON (Unless UF changes their rules regarding COVID).

Masks. UF and The College of Liberal Arts and Sciences emphasizes that, considering CDC recommendations, **face coverings, are expected in all UF facilities.** Regardless of vaccination status, mask usage is a responsible and vital way of preventing transmission of COVID-19. Mask usage is mandatory in offices during one-on-one meetings (students can request a zoom alternative).

Vaccination. If you have not been vaccinated, please consider starting the process immediately. If you are currently at home, you can obtain a first dose where you are and receive a second dose here in Gainesville. If you are already in Gainesville, **you can schedule a vaccination appointment at ONE.UF.** Please keep in mind that being vaccinated for COVID-19 is the best way to protect yourself and others from the potentially life-threatening effects of the virus. Your action in this regard may help ensure the health and safety of yourself, your fellow UF/CLAS students, and the faculty and staff with whom you interact.

Information listed below is subject to change.

Canvas e-learning site: <https://ufl.instructure.com/courses/442263>

All communications must be done through the e-learning site, including homework, deadlines, grades, and announcements. It is your responsibility to check this site for updates. Please do not email the instructors (or the TAs) personal email accounts.

Schedule:

Monday	Period 6	(12:50 PM - 1:40 PM)	LEI0309
Wednesday	Period 6 - 7	(12:50 PM - 2:45 PM)	LEI0309

Professor:

Dr. Adrian E. Roitberg
Office Hours: TBA

Graduate Teaching Assistant:

Jinze Xue (Richard)
Office Hours: TBA

Prerequisites: Some Python programming experience

Textbook:

There is no required textbook.

Course objectives

By the end of this course students should be able to:

1. Handle and visualize different types of datasets common in Chemistry
2. Process data for use in machine learning approaches
3. Identify the suitability of different ML approaches to answer chemical questions based on data
4. Use standard ML python pipelines to train models
5. Assess the quality of the models and their predictive power
6. Recognize different applications of ML in Chemistry

Course Itinerary

(Tentative)

Data handling

Introduction to the course. Python notebooks.

Types of data in biosciences

Chemical data and chemical descriptors

Molecular similarity

RDKit python library for chemical data (<https://github.com/rdkit/rdkit>)

Databases of materials and chemical data

Processing data: analysis and visualization of data

Traditional machine learning

scikit-learn, machine learning in python (<https://scikit-learn.org/stable/>)

Dimensionality reduction: embedding space

Classification: supervised and unsupervised methods

Unsupervised learning: clustering methods

Unsupervised learning: kernel methods and support vector machines (SVM)

Supervised learning: inference models, model training and evaluation

Linear and non-linear classifications models

Decision trees and random forest approaches (RF)

Probabilistic methods

Neural Networks and beyond

Neural networks and deep learning

Processing sequential data: recurrent neural networks (RNN)

Learnable representation, back-propagation

Generative models

Example applications

Boltzmann generators (Noé)

Predicting structural information from data (AlphaFold)

Homework:

There will be homework assigned nearly every week. Homework is usually due one week after it is assigned. It has to be turned in before midnight on the due date. They will be graded and returned. Solutions will be provided after the deadline. Answers should be turned in on time and should be neat and legible.

Exams:

None

Grading: The grade will be determined by Homeworks (50%), Projects (50%).

Grading scale:

A	> 90
A-	87.5 to 89.99
B+	82.5 to 87.49
B	77.5 to 82.49
B-	75 to 77.49
C+	72.5 to 74.99
C	67.5 to 72.49
C-	65 to 67.49
D+	62.5 to 64.99
D	57.5 to 62.49
D-	55 to 57.49
E	<60

Attendance: Lecture attendance is essential for your success in this class. However, we will not take roll. Repeated absence in class will make it very difficult to earn full participation points.

Course Evaluations: Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at <https://evaluations.ufl.edu>. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at <https://evaluations.ufl.edu/results/>.

HONOR CODE

The student honor code can be found at:

<http://www.registrar.ufl.edu/catalog/policies/students.html>

The students, instructor and TAs are honor bound to comply with the Honors Pledge:

We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity.

On all work submitted for credit by students at the university, the following pledge is either required or implied: On my honor, I have neither given nor received unauthorized aid in doing this assignment.

Students with disabilities: Students requiring special accommodations need to register at the Dean of Student Offices and bring the documentation to the instructor.

Counseling services are available at <http://www.counsel.ufl.edu> . or call (352)-392-1575 during regular service hours (8am-5pm). For other hours or weekends call the Alachua County Crisis Center (264-6789). Students may also call the clinician on-call at Student Mental Health for phone callback and consultation at (352)-392-1171

Class Recordings

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third-party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student