Chemistry 213/313: Machine learning in chemistry Spring Quarter 2025

Welcome to Chemistry 213/313! Arguably no area of chemistry has advanced as rapidly as machine learning in the last few years. Machine learning is enabling accurate prediction of molecular properties and reactivities for virtual screening, protein structure prediction, and new ways to interpret complex simulations to name just a few examples; automated chemical synthesis, interpretation of spectra, and sampling molecular conformations are active areas of investigation with the potential to revolutionize molecular and materials design. By the end of this course, you should be able to read and critically evaluate papers on machine learning in chemistry, choose appropriate methods for treating problems, understand their theoretical basis, and implement and train basic models.

Instructor:
Aaron R. Dinner
GCIS E139E, 929 East 57th Street
Email: dinner@uchicago.edu

Teaching assistant:
Kwanghoon Jeong
GCIS E145, 929 East 57th Street
Email: kjeong@uchicago.edu

Lectures:
MWF 10:30-11:20 am
Kent 102

Office hours:
There are no scheduled office hours; you are encouraged to schedule times to meet outside of lecture by email.

Technologies:
To aid in learning, we will be employing academic technologies this quarter. Canvas will be used for course announcements, assignments, distribution of materials, and communication of grades. Some assignments will involve coding, and, particularly for such assignments, the use of large language models (e.g., https://phoenixai.uchicago.edu) is encouraged. You should be sure to understand and critically evaluate their output.

Recommended texts:
Deep Learning for Molecules and Materials by Andrew D. White
Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
Pattern Recognition and Machine Learning by Christopher M. Bishop
Probabilistic Machine Learning: An Introduction by Kevin Murphy
Machine Learning with PyTorch and Scikit-Learn by Sebastian Raschka

Additional resources (unvetted by Prof. Dinner):
Dive into Deep Learning
fast.ai

Reading assignment questions:
Prior to most lectures, a reading assignment will be posted to Canvas. You must do the reading and complete a short assignment. These assignments will be due by 8 am the day of the lecture and will be graded for effort 0, 1, or 2 points total. Generally late responses will not be accepted, though accommodations can be made for religious holidays and documented medical problems with permission of the instructor.

Homework assignments:
Homework assignments will provide hands-on experience with methods. As noted above, large language models are encouraged. You may also work collaboratively. These assignments will also be largely graded on effort. Late homework assignments will be accepted up to 72 hours after the due date and marked down 15% per day.

In-class presentations:
You will be asked to review and lead discussion of two papers of Prof. Dinner's choosing. Depending on enrollment, these may be group presentations.

Exams:
There will be one in-class exam Wednesday, April 30, 2025.

Final projects:
Students will have the opportunity to develop a machine learning project of their choosing. Those enrolled in Chemistry 313 will have an additional final assignment.

Grades:
Grades will be calculated with the following weighting: 15% reading assignment questions, 15% homework assignments, 25% in-class presentations, 15% exam, and 30% final projects.

Expected background and prerequisites:
You should have a general familiarity with quantum mechanics and statistical thermodynamics at the level of typical undergraduate courses and the mathematics required for those courses. Programming experience is helpful but can be gained during the course; we will mainly use Python with PyTorch.

Tentative topics:
We will tentatively be covering the following machine learning topics: classification, regression, dimensional reduction, convolutional neural networks, graph neural networks, transformers, generative models, and reinforcement learning. Illustrative applications in chemistry and related fields will be introduced in parallel.