DATA 22100: Introduction to Machine Learning: Concepts and Applications

Instructor: Dr. Amanda R. Kube Jotte

Winter 2025

E-mail: akube@uchicago.edu

Office: Ryerson 257A

Class Meetings: TTh 2:00pm - 3:20pm

Location: Ryerson Phys Lab 251

Instructor Office Hours: TBD in Ryerson 257A

Course Description

(From Course Listings) This course introduces topics in current applications of machine learning for Data Science minor students. Topics include machine learning models, supervised and unsupervised learning, loss functions, risk, empirical risk and overfitting, regression and classification, clustering, gradient boosting, decision trees and random forests, and a brief introduction to Neural Networks and deep learning.

These topics will be taught and reinforced through lectures, coding activities, labs, homework assignments, exams, and a project.

The only prerequisite for this course is Data 119 which can be taken concurrently. Students who have not taken Data 118 or 119 should have equivalent experience in Python and statistics. You should talk to the instructor if you have questions about this.

Student Learning Objectives

1. Students will be able to analyze and model data in Python using modules such as scikit-learn, statsmodels, and pytorch.
2. Students will be able to explain the difference between supervised and unsupervised learning and know when to apply each.

3. Students will be able to explain the difference between classification and regression and know when to apply each.
4. Students will be able to conduct Maximum Likelihood and Maximum a Posteriori Estimation (MLE and MAP) both mathematically and computationally.
5. Students will be able to explain the function of likelihood and loss functions, identify appropriate loss functions, and calculate likelihood and loss for a model.
6. Students will be able to thoughtfully apply machine learning models including Naive Bayes, Logistic Regression, Random Forests, Gradient Boosting, K-Means Clustering, PCA, and Artificial Neural Networks.
7. Students will be able to apply methods of feature engineering and selection including regularization and to explain when and why these methods are necessary.
8. Students will be able to identify when a model is over- or under-fitting the data and apply methods to combat this.
9. Students will be able to appropriately tune hyperparameters for machine learning models when necessary.
10. Students will be able to interpret the results/output of their models in the context of their data and/or research question.

Course Policies

Course Materials and Announcements

Textbooks You do not need to purchase a printed textbook. Any assigned readings will come from free online textbooks and/or articles and will be posted on Canvas.

Software:
You will need the ability to work with Python scripts and Jupyter Notebooks to complete assignments and view lectures for this course, so you must have access to a computer with Python3 and Jupyter Notebooks installed. You may also install Visual Studio Code and use it to open and edit both .py and .ipynb files. You will also need to install quarto which will be used to convert .ipynb files into pdfs.

Discussion:
We will use Ed and Canvas for all questions and discussions related to the class. Please post questions on Ed rather than sending an email. This serves multiple purposes. First, others may have the same question. Posting to Ed allows us to clarify the issue for everyone at the same time. Second, we are much more likely to respond in a timely manner if you post on Ed as both instructors and TAs will see the post. Lastly, your fellow students may be able to answer your question. One of the best ways to study is to teach the material to someone else [Guerrero and Wiley, 2021, etc].

Before posting to Ed, **please check that no one else has already asked the question**. We will not respond to duplicate questions on Ed. Please view the Ed Discussion Guidelines which you can find in a pinned post on Ed. Using Ed and using it properly allows us to be much more responsive to all students than if we had to answer questions individually.

All announcements related to the class will be made in class, on Canvas, or on Ed. **I will assume that any announcement made on Canvas or on Ed is known to everyone in class within one business day of it being posted.** It is important to check Ed and Canvas regularly! You are responsible for all announcements made in lecture or online.

Emails:
If you have something that you want to talk to me about individually, you are encouraged to send me an email. However, I ask that you please include these things in every email that you send:

1. Your full name as it appears on Canvas
2. The number and/or name of the course you are in
3. The name or number of the assignment you are referring to if applicable

If you do not include these, **I will likely not respond**. Please remember that yours is not the only class that I teach. It saves an incredible amount of time if I do not have to search for your name across all of my courses to figure out who you are and what you need.

Submitting Homework:
Homework assignments will be submitted through Gradescope as a pdf. Many homework assignments include both mathematical and coding questions. There are multiple ways to include both in the same pdf. Some of you may want to get familiar with writing equations using LaTeX. LaTeX math can be written into a markdown cell of a .ipynb file (see this tutorial). You may also find this snipping tool helpful for turning your handwritten math into LaTeX.

Alternatively, you can insert images of handwritten math in an .ipynb file using this syntax in a markdown cell:

```
![alt text](imagename.png "Title")
```

or this syntax in a code cell:

```
from IPython.display import Image
Image(filename='imagename.png',width=800, height=400)
```

which is useful if you need to resize the image.

You can turn your .ipynb file into a pdf using quarto. To do so, you will need to add some YAML to the beginning of your notebook and render the document in your editor of choice. The process for doing this using Jupyter is described here. The process for using Visual Studio Code is described here.

Grading Policy

Your course grade will be calculated as follows:

- Homework - 30%
- Midterm - 30%
- Final Project - 40%

Letter grades will be assigned as follows:

```
def letter_grade(mark):
'''A function to assign letter grades from percentages'''
    if mark >=93:
        grade = 'A'
    elif mark >= 90
        grade = 'A-'
    elif mark >= 87:
        grade = 'B+'
    elif mark >= 83:
        grade = 'B'
    elif mark >= 80:
        grade = 'B-'
    elif mark >= 77:
        grade = 'C+'
    elif mark >= 73:
        grade = 'C'
    elif mark >= 70:
        grade = 'C-'
    elif mark >= 67:
        grade = 'D+'
    elif mark >= 63:
        grade = 'D'
    elif mark >= 60:
        grade = 'D-'
    else:
        grade = 'F'
return grade
```

A Pass/Fail grade may be given upon written request to the instructor before the reading period starts. The grade of P will be awarded only for work of C- or better quality. The grade of Incomplete will be given only in cases of emergency, which will require a conversation with the instructors and Academic Adviser. The grade of W must be requested and discussed with your Academic Adviser by the appropriate deadline designated by the registrar.

Please submit all assignments on time! You will want feedback on your work before completing the next assignments, as most topics in this course build on one another. For this reason, **late**

assignments will not be accepted. I do not allow extensions. However, I welcome you to talk to me if you have issues or if an unusual circumstance arises. This is my policy because I do not think it is my place to choose which reason meets the threshold for an extension.

Because I do not allow late assignments, **I will drop your lowest homework grade.** This is meant to account for any unforeseen issues in the submission process or other problems you may encounter. This includes any reason you may have asked for an extension. It is best if you save this dropped submission for situations like this! I will not drop additional homework grades or grant additional extensions.

Topics to be Covered

Topics we will discuss throughout the quarter will include but are not limited to:

- Parameter Estimation
- Naive Bayes and Gaussians
- Model Evaluation
- Loss Functions and Gradient Decent
- Tree Methods
- k-Means Clustering and Dimensionality Reduction
- Convolutional and Recurrent Neural Networks
- Transformers

A more detailed calendar will be posted to Canvas and updated as needed. This calendar is subject to change.

Use of Generative AI

You will not be allowed to use ChatGPT, Google Bard, or any similar large language models in this class unless I tell you otherwise. Doing so will be treated as a violation of academic integrity. The reason for this, is that it is important you learn to understand and write code on your own in order to be able to properly use such resources in the future.

Academic Integrity

Acting with academic integrity means, in brief, not submitting the statements, work, or ideas of others as one's own. Students are expected to comply with University regulations regarding honest work. If you are in doubt about what constitutes academic dishonesty, speak with me before the assignment is due. Failure to maintain academic integrity on an assignment will result in a penalty befitting the violation, up to and including failing the course and further University sanctions. For more information, consult the student manual https://studentmanual.uchicago.edu/academic-policies/academic-honesty-plagiarism/.

Accommodations

Accessibility: Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please meet with me to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations. For more information, visit disabilities.uchicago.edu.

Accommodations based upon sexual assault: The University is committed to offering reasonable academic accommodations to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If a student comes to us to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if we otherwise observe or become aware of such an allegation, we will keep the information as private as we can, but as faculty members of University of Chicago, we are required to immediately report it to a Department Chair or Dean or directly to the University's Title IX Coordinator. If you would like to speak with the Title IX Coordinator directly, Bridget Collier can be reached at bcollier@uchicago.edu or 773-834-6367. Additionally, you can report incidents or complaints to the Sexual Assault Dean-on-Call (SADoC) by calling 773-834-HELP, or by contacting UCPD at (773)702-8181 or your local law enforcement agency. See https://studentmanual.uchicago.edu/university-policies/the-university-of-chicago-policy-on-title-ix-sexual-harassment/.

First-Generation, Lower-Income and Immigrant Network: As a first-generation, low-income student myself, I understand how these identities can come with additional barriers to success. The University of Chicago FLI Network provides support and community to first-generation, lower-income, and immigrant students and allies. For more information, see: https://flinetwork.uchicago.edu/

Bias Reporting: The University has a process through which students, faculty, staff, and community members who have experienced or witnessed incidents of bias, prejudice or discrimination against a student can report their experiences to the University's Bias Education and Support (BEST) team. See: https://diversityandinclusion.uchicago.edu/resources/reporting-incidents/

Mental Health: Student Wellness' Mental Health professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See: https://wellness.uchicago.edu/mental-health/

Preferred Name and Gender Inclusive Pronouns: In order to affirm each person's gender identity and lived experiences, it is important that we check in with others about pronouns. This simple effort can make a profound difference in a person's experience of safety, respect, and support. See: https://inclusion.uchicago.edu/lgbtq-student-life/lgbtq-resources/

References

[Guerrero and Wiley, 2021] Guerrero, T. A. and Wiley, J. (2021). Expecting to teach affects learning during study of expository texts. *Journal of Educational Psychology*, 113(7):1281.