CMPSC 448: Machine Learning and Algorithmic AI

Penn State University, Fall 2024

Please note that this is a tentative syllabus and subject to change.

Course Information

Lecture

- MWF, 9:05-9:55 am @ Animal, Vet and Biomed Sci 102

Instructor

Rui Zhang

TA

- Sarkar Snigdha Sarathi Das
- Ryo Kamoi

Contact

- For most questions (e.g., homework help, lecture, logistics), please use Piazza
- For personal issues (e.g., exam conflict, honors option), please use Canvas Inbox

Course Goals and Objectives

The goal of this course is to introduce data analysis from the machine learning perspective, in particular how to design and evaluate data-driven solutions for real problems in different domains. Students will gain familiarity with the workings of common machine learning models and will learn how noise and bias in the data affect their results. The course assumes programming skills in Python and knowledge in linear algebra, calculus, basic probability and statistics.

Prerequisites

STAT 318 or STAT 414 and CMPSC 122 or prior programming experience. You are expected to have a good understanding of Linear Algebra, Multivariate Calculus, Probability and Statistics, and Programming Skills. We will cover some background material on these topics early in the lectures. However, it is not meant to replace these regular prerequisite courses. For programming skills, you are expected to feel comfortable processing and analyzing data in Python and be familiar with basic algorithmic design and analysis.

Textbook

The textbook is not required.

- Pattern Recognition and Machine Learning. Christopher Bishop. available online
- Machine Learning: A Probabilistic Perspective. Kevin Murphy
- Elements of Statistical Learning. Hastie, Tibshirani, Friedman. available online
- Reinforcement Learning: An Introduction. Sutton and Barto. available online

Topics and Schedule (Tentative)

We will cover the following topics in four parts.

PART 1: The Basics of Machine Learning and Background
1. Introduction to Machine Learning
2. The Process of Learning and Key Concepts
3. Background I: Linear Algebra and Vector Calculus
4. Background II: Convex Analysis and Optimization
5. Exploratory Data Analysis

PART 2: Supervised Learning
6. Regression: Ordinary Least Squares, Ridge Regression, Principal Component Regression (PCR), and Lasso
7. Nearest Neighbors
8. Artificial Neural Networks (Perceptron & Deep Neural Networks)
9. Logistic Regression
10. Decision Trees
11. Support Vector Machines
12. Ensemble methods: Bagging and Boosting

PART 3: Unsupervised Learning
13. Clustering: k-means, k-means++, and Mixture of Gaussians
14. Principal Component Analysis (PCA)
15. Matrix Factorization

PART 4: Reinforcement Learning
16. Bandits
17. Markov Decision Processes
18. Dynamic Programming
19. Temporal Difference (SARSA and Q-learning)

Grading

- 40%: Homeworks

- 20%: Midterm
- 30%: Final
- 10%: In-Class Quiz

Final letter grades will be curved and the cut scores will be determined after all grades are finalized.

Homeworks and Policy

Late Policy. There is only 1 late day allowed with a penalty of 25% for EACH programming assignment. For example, if you obtain a raw score 90 on the first late day, you will get 90x75%= 67.5; if you submit two days after the deadline, the score is 0.

There will be 5 bi-weekly homeworks, both on theory and programming. Submit your write-up as a single PDF file and Python code on Canvas by 11:59 PM of the due date. The write-up must be neatly typeset as a PDF document. It is strongly recommended to use LaTex that can produce mathematical symbols and equations with high legibility. In addition, please include the following at the top of the first page: your name, your PSU ID, and the ID of any students with whom you discussed the homework.

You are welcome and encouraged to discuss homework with other students (2 or 3 students) in a small group, along with other course materials. However, the write-up/code must be written by yourself, and you must list in the write-up all student IDs with whom you discussed the homework. The collaboration may include brainstorming and verbal discussion of possible solutions, but must not go as far as directly telling how to solve a problem. You may not look at other student's write-up/code, whether partial or complete.

Notes on Python: The End of Life for Python 2 has already been announced, and programming shall be done with Python 3. We recommend the Anaconda distribution of Python 3 coming with many useful packages installed (NumPy, SciPy, pandas, scikit-learn, Matplotlib) and allowing easy installation of TensorFlow and PyTorch.

Academic Integrity

Academic integrity is the pursuit of scholarly activity in an open, honest and responsible manner. Academic integrity is a basic guiding principle for all academic activity at The Pennsylvania State University, and all members of the University community are expected to act in accordance with this principle. Consistent with this expectation, the University’s Code of Conduct states that all students should act with personal integrity, respect other students’ dignity, rights and property, and help create and maintain an environment in which all can succeed through the fruits of their efforts.Academic integrity includes a commitment by all members of the University community not to engage in or tolerate acts of falsification,

misrepresentation or deception. Such acts of dishonesty violate the fundamental ethical principles of the University community and compromise the worth of work completed by others.

The Department of Computer Science and Engineering expects all student programming work assigned in a class to be completed independently by students (or by teams if permitted/required) and to consist of code designed and developed solely by the students. The use of any other code is not permitted unless the course instructor explicitly allows it and such code is clearly identified as coming from an external source and that source is credited. Students will never be given credit for code which they did not construct. Department policy for academic sanctions of academic integrity violations specifies a reduction of score on the submission (typically reduced to a 0 except for minor infractions) and a reduction of up to 1 letter grade for the final course grade. For students with previous academic integrity violations (occurring in any course), the department will recommend to the College of Engineering Academic Integrity Committee that the student receive an F in the course. For more information, please check Academic Integrity Standards for CMPSC, CMPEN, and CSE Programming Courses.

Accommodations for Students with Disabilities

Penn State welcomes students with disabilities into the University's educational programs. Every Penn State campus has an office for students with disabilities. Student Disability Resources (SDR) website provides contact information for every Penn State campus (http://equity.psu.edu/sdr/disability-coordinator). For further information, please visit Student Disability Resources website (http://equity.psu.edu/sdr/).

In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: See documentation guidelines (http://equity.psu.edu/sdr/guidelines). If the documentation supports your request for reasonable accommodations, your campus disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early as possible. You must follow this process for every semester that you request accommodations.

Reporting Educational Equity Concerns through the Report Bias Website

Consistent with University Policy AD29 Statement on Intolerance (https://policy.psu.edu/policies/ad29) , students who believe they have experienced or observed a hate crime, an act of intolerance, discrimination, or harassment that occurs at Penn State are

urged to report these incidents as outlined on the University's Report Bias webpage (http://equity.psu.edu/reportbias/)

Counseling & Psychological Services

Many students at Penn State face personal challenges or have psychological needs that may interfere with their academic progress, social development, or emotional wellbeing. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients' cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

- Counseling and Psychological Services at University Park (CAPS (http://studentaffairs.psu.edu/counseling/): 814-863-0395
- Counseling and Psychological Services at Commonwealth Campuses (https://senate.psu.edu/faculty/counseling-services-at-commonwealth-campuses/)
- Penn State Crisis Line (24 hours/7 days/week): 877-229-6400
  Crisis Text Line (24 hours/7 days/week): Text LIONS to 741741

Frequently Asked Questions

Q: How is this course different from other applied or data science courses?
A: Data science is a broad term for multiple disciplines, machine learning fits within data science. While the focus of data science is mainly on skills and tools (e.g., Spark, MapReduce, etc) that are required to tackle big data including feature engineering, data cleaning, preparation, exploratory data analysis, and utilizing ML algorithms, the emphasis of this course will be on machine learning algorithms and applications, with some broad explanation of the underlying principles. However, several software libraries and data sets publicly available will be used to illustrate the application of these algorithms (e.g., pandas for loading and exploratory data analysis and scikit-learn for algorithms). In your homework assignments, you will also be asked to design and implement new learning algorithms from scratch.

Q: How much will the workload be?
A: This is a 3-credit course. You can expect you will spend on average 10-15 hours per week for this course including attending lectures, doing homework, studying for exams, etc. In particular, proficiency in linear algebra and probability is strongly recommended. Otherwise, you would spend more time.