PHYS 2200: *Applied Data Science*
Fall 2024

Instructor

Prof. Masao Sako
Office: DRL 4N2
email: masao@sas.upenn.edu
phone: 215-898-8151

Course Description

This is the second of a two-semester gateway course on programming, data analysis, and data science in Python. This semester will focus on big data, machine learning, and artificial intelligence and we will dive deeper into the practical applications of these data science methodologies using real-world data. Topics covered include supervised and unsupervised machine learning, decision trees, random forests, neural networks, and deep learning. Some modern methods such as transformers and generative AI will also be discussed. Finally, we will explore effective ways of using AI chatbots such as ChatGPT for efficiently building software.

Pre-requisite

- PHYS1100 or equivalent
- MATH1400 (with concurrence)
- Proficiency in Python and basic calculus

Primary Textbook

- *Hands on Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*, Aurelien Geron, 3rd Edition, O'Reilly Media, (ISBN: 978-1098125974)
- *Deep Learning with Python*, Francois Chollet, 2nd Edition, Manning, (ISBN: 978-1617296864)

Course Structure

This course is a practical course on analyzing datasets using machine learning and artificial intelligence. The course is 1.5 cu and will meet 6 hours/week — 3 hours of lectures plus 3 hours of coding recitation in small groups of 10-15 students. Lectures will cover the theory and background while the recitations will be hands on and cover the practical aspects of machine learning and deep learning methods using modern Python libraries and modules.

Assessment

- **50% weekly assignments**
    - Assignments will mainly consist of coding exercises and open-ended questions. Students will turn them in as Jupyter notebooks through Canvas.

- 20% **coding recitation participation**
  - These hands-on labs are an essential part of the course where students will build their technical skills.
  - Attendance and participation is required. However, the activities will be graded based only on completion.

- 10% **biweekly quizzes**
  - Short in-class quizzes to ensure that students are keeping pace with the course.

- 20% **final project**
  - For the final project, students will analyze a real-world dataset using AI and put together a full analysis report.